THE ANNALS *of* MATHEMATICAL STATISTICS

[illegible] AND EDITED BY [illegible]

EDITED BY [illegible]

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. [illegible], No. 2

THE [illegible]
OF MATHEMA[illegible] STATISTICS

[illegible]

T. W. A[illegible]

Associate Editors

R. C. BOSE	M. A. [illegible]	[illegible]ANDER M. MOOD
[illegible] FELLER	E. L. LEHMANN	[illegible] W. TUKEY

with the cooperation of

M. [illegible]	J. L. Hodges, [illegible]	J. Neyman
[illegible] W. [illegible]	[illegible]	Edward Paulson
David [illegible]	[illegible]	H. [illegible]
[illegible]	Harold [illegible]	[illegible]
William G. Cochran	Howard [illegible]	[illegible]
[illegible] Daly	[illegible]	[illegible]
[illegible] L. Doob	H. [illegible]	[illegible]
[illegible] Harris	[illegible]	[illegible]

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

INSTITUTE OF MATHEMATICAL STATISTICS

[illegible] To June 30, 195[illegible]

[illegible] Business Administration Building, University of Michigan, Ann Arbor, Michigan. [illegible]

[illegible] From July 1, 195[illegible]

[illegible] of [illegible], Michigan State College, East Lansing, Michigan. [illegible]

This address should be used for all communications concerning membership, subscriptions, changes of address, back numbers, etc., but not for editorial correspondence. Changes in mailing address which are to become effective for a given issue should be reported to the [illegible] on or before the 15th of the month preceding the month of that issue.

[illegible] Department of Mathematical Statistics, Columbia University, New York 27, New York. [illegible] W. Anderson, Editor.

[illegible] to this address until [illegible] 21, 195[illegible]. All other editorial correspondence should be directed to this address [illegible] year 19[illegible]. Manuscripts submitted for [illegible] after August 1, 19[illegible] should be sent to [illegible] E. L. Lehmann, Statistical Laboratory, University of California, Berkeley 4, California.

[illegible] $12 per year in [illegible] and Canada. $[illegible] elsewhere. [illegible] $3.00. Back numbers [illegible] $[illegible] per volume or $3.[illegible] per issue.

[illegible]

[illegible] Maryland, U. S. A.

[illegible] March 3, 1879.

THE LARGE-SAMPLE POWER OF TESTS BASED ON PERMUTATIONS OF OBSERVATIONS[1]

By Wassily Hoeffding
University of North Carolina

Summary. The paper investigates the power of a family of nonparametric tests which includes those known as tests based on permutations of observations. Under general conditions the tests are found to be asymptotically (as the sample size tends to ∞) as powerful as certain related standard parametric tests. The results are based on a study of the convergence in probability of certain random distribution functions. A more detailed summary will be found at the end of the Introduction.

1. Introduction. Let X be a random variable whose values are points x in a space $\mathfrak{X}$. The probability distribution of X is characterised by the probability measure $P(A) = \Pr\{X \varepsilon A\}$, defined on an additive class $\mathfrak{A}$ of subsets A of $\mathfrak{X}$. (In the applications to be considered $\mathfrak{X}$ can be taken as a finite-dimensional Euclidean space, $\mathfrak{A}$ as the family of Borel sets.) Let $\mathfrak{G}$ be a finite group of transformations g of $\mathfrak{X}$ onto itself which also map $\mathfrak{A}$ onto itself. Thus, for every g in $\mathfrak{G}$, every x in $\mathfrak{X}$ and every A in $\mathfrak{A}$, the point gx is in $\mathfrak{X}$ and the set gA of points gx, $x \varepsilon A$, is in $\mathfrak{A}$. Let M be the number of elements in $\mathfrak{G}$. Let H be a hypothesis which implies that the distribution of X is invariant under the transformations in $\mathfrak{G}$, so that for every g in $\mathfrak{G}$, gX has the same distribution as X.

For example, let $\mathfrak{X}$ be the n-dimensional Euclidean space, and let H be the hypothesis that the components $X_1, \cdots, X_n$ of X are independent, each X_i being symmetrically distributed about the median 0. Then H implies that the distribution of X is invariant under changes of sign of the X_i. Here $M = 2^n$. Alternatively, if $X_1, \cdots, X_m, \cdots, X_n$ are independent, $X_1, \cdots, X_m$ have a common distribution and $X_{m+1}, \cdots, X_n$ have a common distribution, then the distribution of X is invariant under the $M = m!\,(n-m)!$ permutations which permute the first m or the last $n - m$ components.

All real-valued functions of x to be considered are understood to be measurable $(\mathfrak{A})$. The expected value of a function $f(X)$ when X has distribution P will be denoted by $E_P f(X)$ or $Ef(X)$.

By a test of H we shall mean a function $\phi(x)$, $0 \leqq \phi(x) \leqq 1$, which expresses the probability with which H is rejected when X takes the value x. The power of the test ϕ with respect to P (the unconditional probability of rejecting H when P is the true distribution and test ϕ is used) is equal to $E_P\phi(X)$. If $E_P\phi(X) = \alpha$ whenever H is true, the test ϕ is said to be similar of size α for testing H.

This paper will be mainly concerned with tests of the following type. Let $t(x)$ be a real-valued function on $\mathfrak{X}$. For every $x \varepsilon \mathfrak{X}$ let

$$t^{(1)}(x) \leqq t^{(2)}(x) \leqq \cdots \leqq t^{(M)}(x)$$

[1] Work done under the sponsorship of the Office of Naval Research.

be the ordered values $t(gx)$, for all g in $\mathcal{G}$. Given a number α, $0 < \alpha < 1$, let k be defined by

$$k = M - [M\alpha],$$

where $[M\alpha]$ denotes the largest integer less than or equal to $M\alpha$. Let $M^+(x)$ and $M^0(x)$ be the numbers of values $t^{(j)}(x)$, $(j = 1, \cdots, M)$, which are greater than $t^{(k)}(x)$ and equal to $t^{(k)}(x)$, respectively, and let

$$a(x) = \frac{M\alpha - M^+(x)}{M^0(x)}.$$

Since $M^+(x) \leqq M - k \leqq M\alpha$ and $M^+(x) + M^0(x) \geqq M - k + 1 > M\alpha$, we have $0 \leqq a(x) < 1$.

Let the test $\phi(x)$ be defined by

$$\phi(x) = \begin{cases} 1 & \text{if} \quad t(x) > t^{(k)}(x), \\ a(x) & \text{if} \quad t(x) = t^{(k)}(x), \\ 0 & \text{if} \quad t(x) < t^{(k)}(x). \end{cases} \tag{1.1}$$

For every $x \, \varepsilon \, \mathfrak{X}$ we have

$$\textstyle\sum_g \phi(gx) = M^+(x) + a(x)M^0(x) = M\alpha,$$

where $\sum_g$ stands for summation over all g in $\mathcal{G}$. If the distribution P of X is invariant under all g in $\mathcal{G}$, we have

$$M\alpha = E_P \textstyle\sum_g \phi(gX) = \sum_g E_P \phi(X) = ME_P \phi(X).$$

Hence the test ϕ is similar of size α for testing H.

Tests which are essentially of the form (1.1) have been considered by R. A. Fisher [3], Pitman [11], Welch [14]. Lehmann and Stein [8] have shown that tests of this type, with suitable functions $t(x)$, are most powerful (or most powerful similar, etc.) for testing certain nonparametric hypotheses H against specified alternatives.

A test of the form (1.1) differs from a conventional test mainly in that the "critical value," $t^{(k)}(X)$, is a random variable. This circumstance makes the exact evaluation of its power function difficult. It will, however, be shown that under certain conditions $t^{(k)}(X)$ is close to a constant with high probability. Then the power of the test can be approximated in terms of the distribution function of $t(X)$.

More precisely, suppose that the objects so far considered, $\mathfrak{X} = \mathfrak{X}_n$, $\mathcal{G} = \mathcal{G}_n$, $t(x) = t_n(x)$, etc., are defined for an infinite sequence of positive integers n. It will be assumed that the size α of the test is fixed and that $M \to \infty$ as $n \to \infty$. Then

$$k/M \to 1 - \alpha \qquad \text{as } n \to \infty.$$

Suppose that for a given sequence $\{P_n\}$ of distributions of $X = X^{(n)}$ the following two conditions are satisfied:

CONDITION A. *There exists a constant* λ *such that* $t_n^{(k)}(X) \to \lambda$ *in probability.*

CONDITION B. *There exists a function* $H(y)$, *continuous at* $y = \lambda$, *such that for every* y *at which* $H(y)$ *is continuous*

$$\Pr\{t_n(X) \leqq y\} \to H(y).$$

From (1.1) we have

$$\Pr\{t_n(X) > t_n^{(k)}(X)\} \leqq E_{P_n}\phi_n(X) \leqq \Pr\{t_n(X) \geqq t_n^{(k)}(X)\}. \tag{1.2}$$

Hence it follows that Conditions A, B imply

$$E_{P_n}\phi_n(X) \to 1 - H(\lambda). \tag{1.3}$$

It should be noted that the function $t(x)$ in the definition (1.1) of $\phi(x)$ can be replaced by any function $t'(x)$ such that for every x in $\mathfrak{X}$ and every two elements g, g' of $\mathcal{G}$ the difference $t'(gx) - t'(g'x)$ has the same sign as $t(gx) - t(g'x)$. For example, this is true for $t'(x) = c(x)f(t(x)) + d(x)$, where $f(y)$ is an increasing function, $c(x) > 0$, and $c(x)$, $d(x)$ are invariant under $\mathcal{G}$ (cf. Lehmann and Stein [8]). Thus if Conditions A, B are not satisfied, they may possibly be satisfied after $t_n(x)$ has been replaced by a suitable function $t_n'(x)$.

In general λ and $H(y)$ will depend on the sequence $\{P_n\}$. It will, however, be seen that the dependence of λ on $\{P_n\}$ is much less pronounced than that of $H(y)$, in the sense that for a class C of sequences $\{P_n\}$ the value λ is the same while $1 - H(\lambda)$ ranges from α to 1.

For every x in $\mathfrak{X}$ let $MF_n(y, x)$ be the number of elements g in $\mathcal{G}$ for which $t_n(gx) \leqq y$. For x fixed, $F_n(y, x)$ is a distribution function. Suppose that for some sequence $\{P_n\}$ the following condition is satisfied:

CONDITION A′. $F_n(y, X) \to F(y)$ *in probability for every* y *at which* $F(y)$ *is continuous, where* $F(y)$ *is a distribution function, the equation* $F(y) = 1 - \alpha$ *has a unique solution* $y = \lambda$, *and* $F(y)$ *is continuous at* $y = \lambda$.

It will be shown in Section 3 that A′ implies that $t_n^{(k)}(X) \to \lambda$ in probability, so that A is satisfied with λ as defined in A′; furthermore, if H is true for every P_n of the sequence, $t_n(X)$ has the limiting distribution function $F(y)$.

Let ϕ_n^* be a test of the conventional form $\phi_n^*(x) = 1$, a_n^*, or 0 according as $t_n(x) > \lambda_n$, $= \lambda_n$, or $< \lambda_n$, where $0 \leqq a_n^* \leqq 1$ and λ_n is a constant. Suppose that λ_n and a_n^* are so chosen that the test ϕ_n^* has size α for testing that $P_n = P_n^*$, a distribution for which H is true. It follows from the preceding paragraph that if A′ is satisfied for $\{P_n^*\}$, then $\lambda_n \to \lambda$. Moreover, if B holds,

$$E_{P_n}\phi_n^*(X) \to 1 - H(\lambda). \tag{1.4}$$

Hence if $C(\lambda)$ denotes the class of all sequences $\{P_n\}$ for which A′, with λ fixed, and B, with some $H(y)$, are satisfied, and if $C(\lambda)$ contains $\{P_n^*\}$, then the powers of the tests ϕ_n and ϕ_n^* tend to the same limit for every $\{P_n\}$ in $C(\lambda)$. The nonparametric test ϕ_n can be said to be asymptotically as powerful with respect to $C(\lambda)$ as ϕ_n^*. This result will be of particular interest when ϕ_n^* is a most powerful, or otherwise "optimum," parametric test, as in the examples of this paper.

It also can happen that for different sequences $\{P_n\}$, $t^{(k)}(X)$ converges to different values λ, but in every case the test ϕ_n is asymptotically as powerful as the most powerful test for a parametric family of distributions to which P_n belongs. This point will be illustrated in Section 7.

In most applications to be considered, $H(y)$ is either a (cumulative) distribution function, or $H(y) \equiv 0$. In the latter case the relations (1.3) and (1.4) merely imply that both tests are consistent (have limiting power 1). The case $0 < H(\lambda) < 1$ will usually occur when P_n approaches, in a certain sense, the null hypothesis. For example, let P_n be the distribution of two independent random samples of m and $n - m$ observations from two normal distributions with means $\mu_1 \leqq \mu_2$ and common variance σ^2. Let $\mathcal{G}$ consist of the $M = n!$ permutations of the n observations. Let $t_n(x)$ be the standard t-statistic for two samples. The results of Section 6 imply that Condition A′ is satisfied with $F(y) = \Phi(y)$, where

$$\Phi(y) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} e^{-\frac{1}{2}t^2}\, dt. \tag{1.5}$$

Condition B is satisfied with $H(y) = \Phi(y - c)$ if $(\mu_2 - \mu_1)\, \sigma^{-1} \cdot \{m(n - m)/n\}^{\frac{1}{2}}$ tends to a finite limit c. This will not be the case if, as is frequently assumed, $m/n \to p$, $0 < p < 1$, and $\delta = (\mu_2 - \mu_1)\sigma^{-1}$ is independent of n. In this case one can, however, conclude that if δ is sufficiently small, the number N of observations required to achieve the power $1 - \Phi(\lambda - c)$ is approximately given by $\delta\{p(1 - p)N\}^{\frac{1}{2}} = c$, and this is true for either test. In this sense the asymptotic relative efficiency of the two tests is arbitrarily close to one for δ sufficiently small.

The main object of this paper is to indicate several methods for ascertaining that Condition A is satisfied. By way of illustration the methods are applied to a number of tests which have been considered in the literature. In Section 2 bounds for $t^{(k)}(x)$ are obtained which provide a simple criterion for consistency. Sufficient conditions for the convergence to zero of the variance of the random variable $F_n(y, X)$ are given (Section 3) and used to obtain the large-sample power of several tests (Sections 4–7). The remaining Sections 8–10 show how a theorem can be applied which gives sufficient conditions for the convergence of $F_n(y, x^{(n)})$, for a sequence of fixed values $x^{(n)}$. The fulfilment of these conditions in probability for a sequence of random variables $X^{(n)}$ is found to be sufficient for the convergence in probability of $F_n(y, X^{(n)})$. An extension to random distributions of the second limit theorem of probability theory (Section 10) generalizes a recent result of Ghosh [6].

2. Bounds for $t^k(x)$; consistency. In this section it will be shown that, given a test $\phi(x)$ of the form (1.1), the function $t(x)$ can always be so chosen that one or two moments of the distribution function $F_n(y, x)$ are (essentially) fixed for all x, and the critical value $t^{(k)}(x)$ is confined to a finite interval which depends only on α.

Let G be a random variable whose values are the M elements g of $\mathcal{G}$, each element having the same probability M^{-1}. Then $F_n(y, x)$, as defined in Section 1, is the distribution function of the random variable $t(Gx)$.

Let $m(x)$ and $v(x)$ denote the mean and the variance of $t(Gx)$, so that

$$m(x) = M^{-1} \sum_g t(gx), \qquad v(x) = M^{-1} \sum_g [t(gx) - m(x)]^2.$$

Let $t'(x) = v(x)^{-\frac{1}{2}}[t(x) - m(x)]$ if $v(x) > 0$, $t'(x) = 0$ if $v(x) = 0$. Then the test $\phi(x)$ in (1.1) is not changed if $t(x)$ is replaced by $t'(x)$. Thus we may always assume that the distribution function $F_n(y, x)$ has mean 0 and variance less than or equal to 1. If a probability limit $F(y)$ of $F_n(y, X)$ exists for all y, then $F(y)$ is a distribution function with the same properties. If, moreover, the probability of $t(gX) = t(X)$ for all g in $\mathcal{G}$ tends to 0 as $n \to \infty$, then the probability of $v(X) = 0$ tends to 0, and $F(y)$ has variance 1. In a similar way, if $t(x) \geqq 0$, we may, for instance, replace $t(x)$ by a function $t'(x)$ such that $t'(x) \geqq 0$ and $Et'(Gx) = c$, an arbitrary positive constant.

THEOREM 2.1. *If* $t(x) \geqq 0$, $Et(Gx) = c > 0$, *then*

$$t^{(k)}(x) < \frac{c}{\alpha}. \tag{2.1}$$

If $Et(Gx) = 0$, $Et(Gx)^2 \leqq 1$, *then*

$$-\left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{2}} \leqq t^{(k)}(x) < \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{2}}. \tag{2.2}$$

PROOF. We have

$$MF(t^{(k)}(x) - 0, x) \leqq k - 1 < M - M\alpha \leqq k \leqq MF_n(t^{(k)}(x), x),$$

so that

$$F_n(t^{(k)}(x) - 0, x) < 1 - \alpha \leqq F_n(t^{(k)}(x), x).$$

If $t(x) \geqq 0$, $Et(Gx) = c$, then for every $z > 0$

$$1 - F_n(z - 0, x) = \Pr\{t(Gx) \geqq z\} \leqq \frac{c}{z}.$$

Hence (2.1).

If $Et(Gx) = 0$, $Et(Gx)^2 = c^2 \leqq 1$, relation (2.2) follows in a similar way by using the inequalities of Tchebycheff-Cantelli (see, e.g., [4], p. 126 or [12], p. 198)

$$F_n(y, x) \leqq \frac{1}{1 + c^{-2}y^2}, \qquad \text{if } y < 0,$$

$$F_n(y - 0, x) \geqq 1 - \frac{1}{1 + c^{-2}y^2} \qquad \text{if } y > 0.$$

Apart from providing, via (1.2), crude bounds for the power of ϕ, Theorem 2.1 permits us to draw the following conclusion. If $t_n(x)$ satisfies either of the conditions of the theorem and, for some sequence $\{P_n\}$ of distributions, $H(y) = \lim \Pr\{t_n(X) \leqq y\} = 0$ for all real y, which is a sufficient condition for consistency of the tests ϕ_n^*, then the tests ϕ_n are also consistent. This result is independent of whether $t_n^{(k)}(X)$ converges in probability to a constant.

3. Sufficient conditions for the convergence in probability of $t_n^{(k)}(X)$.

THEOREM 3.1. *Suppose that for a sequence* $\{P_n\}$ *of distributions of* $X = X^{(n)}$, $F_n(y, X)$ *tends in probability to* $F(y)$ *for every* y *at which* $F(y)$ *is continuous, where* $F(y)$ *is a distribution function and the equation* $F(y) = 1 - \alpha$ *has a unique solution* $y = \lambda$. *Then* $t_n^{(k)}(X) \to \lambda$ *in probability.*

PROOF. By the definitions of $t_n^{(k)}(x)$ and $F_n(y, x)$,

$$\Pr\{t_n^{(k)}(X) \leqq y\} = \Pr\{F_n(y, X) \geqq k/M\} \tag{3.1}$$

for every real y. Let y be a point of continuity of $F(y)$. Since, by assumption, $k/M \to 1 - \alpha = F(\lambda)$, and $y < \lambda$ implies $F(y) < F(\lambda)$, the right-hand side of (3.1) tends to 0 if $y < \lambda$. Similarly it tends to 1 if $y > \lambda$. Hence $t_n^{(k)}(X) \to \lambda$ in probability.

A sufficient condition for a sequence of random variables to converge in probability to a constant c is that their means and variances converge, respectively, to c and 0. If the random variables are uniformly bounded, the condition is also necessary. Hence $F_n(y, X) \to F(y)$ in probability if and only if

$$EF_n(y, X) \to F(y), \qquad EF_n(y, X)^2 \to F(y)^2. \tag{3.2}$$

We can write

$$F_n(y, x) = M^{-1} \textstyle\sum_g C(gx),$$

where $C(x) = 1$ or 0 according as $t_n(x) \leqq y$ or $>y$. Hence

$$EF_n(y, X) = M^{-1} \textstyle\sum_g \Pr\{t_n(gX) \leqq y\}, \tag{3.3}$$

$$EF_n(y, X)^2 = M^{-2} \textstyle\sum_g \sum_{g'} \Pr\{t_n(gX) \leqq y,\ t_n(g'X) \leqq y\}. \tag{3.4}$$

Let G be the random transformation defined in Section 2, let G' have the same distribution as G, and let G, G' and X be mutually independent. Then equations (3.3), (3.4) can be written as

$$EF_n(y, X) = \Pr\{t_n(GX) \leqq y\}, \tag{3.5}$$

$$EF_n(y, X)^2 = \Pr\{t_n(GX) \leqq y,\ t_n(G'X) \leqq y\}. \tag{3.6}$$

Note that $t_n(GX)$ and $t_n(G'X)$ are identically distributed, but not independent (except in the trivial case when the random variable $F_n(y, X)$ has variance 0). Equations (3.5) and (3.6) imply that (3.2) is satisfied if $t_n(GX)$ has the limiting distribution function $F(y)$, and $t_n(GX)$ and $t_n(G'X)$ are independent in the limit. Making use of Theorem 3.1, we can state

THEOREM 3.2. *Suppose that, for some sequence* $\{P_n\}$ *of distributions,* $t_n(GX)$ *and* $t_n(G'X)$ *have the limiting joint distribution function* $F(y)F(y')$. *Then for every* y *at which* $F(y)$ *is continuous*

$$F_n(y, X) \to F(y) \textit{ in probability,}$$

and if the equation $F(y) = 1 - \alpha$ *has a unique solution* $y = \lambda$,

$$t_n^{(k)}(X) \to \lambda \textit{ in probability.}$$

We also observe the following. If H is true, $t_n(GX)$ and $t_n(X)$ have the same distribution. Thus if $F_n(y, X) \to F(y)$ in probability for a sequence of distributions invariant under $\mathcal{G}$, then $t_n(X)$ has the limiting distribution $F(y)$. An implication concerning the test ϕ_n^* was pointed out in the Introduction.

The next theorem, 3.3, gives conditions under which two functions $t_n(x)$ and $t'_n(x)$ are, in a certain sense, asymptotically equivalent.

THEOREM 3.3. *Let* $t'_n(x) = c_n(x)t_n(x) + d_n(x)$, *where*

$$c_n(GX) \to 1 \text{ and } d_n(GX) \to 0 \text{ in probability,} \tag{3.7}$$

and let $F'_n(y, x) = \Pr\{t'_n(Gx) \leq y\}$. *Then*

$$F_n(y, X) \to F(y) \text{ in probability} \tag{3.8}$$

if and only if

$$F'_n(y, X) \to F(y) \text{ in probability.} \tag{3.9}$$

PROOF. It is sufficient to show that (3.8) implies (3.9). As has been seen, (3.8) is equivalent to $\Pr\{t_n(GX) \leq y\} \to F(y)$, $\Pr\{t_n(GX) \leq y, t_n(G'X) \leq y\} \to F(y)^2$. Due to assumption (3.7) these relations remain true if $t_n(x)$ is replaced by $t'_n(x)$. This implies (3.9).

The fulfilment of the conditions of Thorem 3.2 can frequently be demonstrated with the aid of the central limit theorem for vectors. One version of this theorem, which will be of particular use in Section 6, is stated below as Theorem 3A. It easily follows from Uspensky's proof [12] of the central limit theorem for vectors.

THEOREM 3A. *Let* $(Y_1, Y'_1), (Y_2, Y'_2), \cdots, (Y_n, Y'_n)$ *be n independent random vectors,* $EY_i = EY'_i = 0$, $E|Y_i|^3 < \infty$, $E|Y'_i|^3 < \infty$. *Let*

$$\bar{Y} = \sum_1^n Y_i\left(\sum_1^n EY_i^2\right)^{-\frac{1}{2}}, \qquad \bar{Y}' = \sum_1^n Y'_i\left(\sum_1^n EY_i'^2\right)^{-\frac{1}{2}},$$

$$\rho = E\bar{Y}\bar{Y}',$$

$$\omega = \sum_1^n E|Y_i|^3\left(\sum_1^n EY_i^2\right)^{-3/2}, \qquad \omega' = \sum_1^n E|Y'_i|^3\left(\sum_1^n EY_i'^2\right)^{-3/2}.$$

Then for any two real numbers y, y'

$$|\Pr\{\bar{Y} \leq y, \bar{Y}' \leq y'\} - \Phi(y)\Phi(y')| \leq f(\rho, \omega, \omega'),$$

where $\Phi(y)$ *is defined by* (1.5) *and the function* $f(u, v, w)$ *is independent of* n, y, y' *and of the distribution of the* Y_i, Y'_i, *and* $f(u, v, w) \to 0$ *as* $u \to 0, v \to 0, w \to 0$.

4. Test for the median of a symmetrical distribution. Let $\mathfrak{X}$ be the Euclidean n-dimensional space and H the hypothesis that the components $X_1, \cdots, X_n$ of the random vector X are independent and each X_i is symmetrically distributed about the median 0. H implies that the distribution of X is invariant under the $M = 2^n$ transformations $gX = ((-1)^{j_1}X_1, \cdots, (-1)^{j_n}X_n)$, $j_i = 0$ or 1, $i = 1$,

$\cdots, n$. The random transformation Gx of x can be written $Gx = (G_1x_1, \cdots, G_nx_n)$, where $G_1, \cdots, G_n$ are independent, $G_i = -1$ or 1 with probabilities $\frac{1}{2}, \frac{1}{2}$. Let $\phi(x)$ be the test (1.1) with

$$t(x) = \sum_1^n x_i \left(\sum_1^n x_i^2\right)^{-\frac{1}{2}},$$

or $t(x) = 0$ if $\sum_1^n x_i^2 = 0$. The factor $(\sum x_i^2)^{-\frac{1}{2}}$ is invariant under the transformations g and is so chosen that $t(Gx)$ has mean 0 and variance 1 (unless $x_1 = \cdots = x_n = 0$). Bounds for $t^{(k)}(x)$ can be obtained from Theorem 2.1.

It follows from the results of Lehmann and Stein [8] that the test ϕ is most powerful similar for testing H against the alternative that $X_1, \cdots, X_n$ are independent with a common normal distribution whose mean is positive; the test ϕ with $t(x)$ replaced by $|t(x)|$ is most stringent similar for testing H against the alternative of a common normal distribution with nonzero mean. It will suffice to consider the former, "one-sided" test. The results will be easily applicable to the "two-sided" case.

Let $Y_i = G_iX_i$, $Y'_i = G'_iX_i$, where all G_i, G'_i are independent, identically distributed, and independent of the X_i. Then $Y_i^2 = Y_i'^2 = X_i^2$,

$$t(GX) = n^{-\frac{1}{2}} \sum_{i=1}^n Y_i \left(n^{-1} \sum_{i=1}^n X_i^2\right)^{-\frac{1}{2}},$$

$$t(G'X) = n^{-\frac{1}{2}} \sum_{i=1}^n Y'_i \left(n^{-1} \sum_{i=1}^n X_i^2\right)^{-\frac{1}{2}}.$$

Suppose that $X_1, \cdots, X_n$ are independent and identically distributed with mean μ and positive variance σ^2. By Khintchine's theorem,

$$n^{-1} \sum_1^n X_i^2 \to \sigma^2 + \mu^2$$

in probability. Hence $(t(GX), t(G'X))$ has the same limiting distribution (if any) as

$$\left((\sigma^2 + \mu^2)^{-\frac{1}{2}} n^{-\frac{1}{2}} \sum_1^n Y_i, \quad (\sigma^2 + \mu^2)^{-\frac{1}{2}} n^{-\frac{1}{2}} \sum_1^n Y'_i\right). \tag{4.1}$$

The vectors $(Y_1, Y'_1), \cdots, (Y_n, Y'_n)$ are independent and identically distributed, with

$$EY_i = EY'_i = 0, EY_i^2 = EY_i'^2 = \sigma^2 + \mu^2, EY_iY'_i = EG_iG'_iX_i^2 = EG_iEG'_iEX_i^2 = 0.$$

By the central limit theorem for identically distributed vectors (see, e.g., Cramér [2], p. 286), the random vector (4.1) has the limiting distribution function $\Phi(y)\Phi(y')$. The same is true of $(t(GX), t(G'X))$. By Theorem 3.2, $t^{(k)}(X) \to \lambda$ in probability, where $\Phi(\lambda) = 1 - \alpha$.

Under the same conditions we have for every fixed y

$$\lim_{n\to\infty} \Pr\{t(X) \leq (y + n^{\frac{1}{2}}\mu/\sigma)(1 + (\mu/\sigma)^2)^{-\frac{1}{2}}\} = \Phi(y).$$

Hence if μ/σ is independent of n (as is implied in the assumptions) and positive, the function $H(y)$ of Section 1 is $\equiv 0$, and the power of the test tends to 1. It follows from the Lyapunov form of the central limit theorem and its extension to vectors (for example, Theorem 3A) that all results remain true if the common distribution of $X_1, \cdots, X_n$ depends on n, provided $E\,|\,X_1\,|^3\sigma^{-3} = o(n^{\frac{1}{2}})$. If $(\mu/\sigma)n^{\frac{1}{2}}$ converges to a constant c, then $H(y) = \Phi(y - c)$. An alternative interpretation of this result, with μ/σ fixed but small, is indicated in the Introduction.

The function $t(x)$ is an increasing function of Student's statistic for testing whether the mean of n independent random variables with a common normal distribution is zero. Thus the test ϕ_n^* of Section 1, with suitably chosen λ_n, is equivalent to Student's (one-sided) test whose size (for testing the normal hypothesis) is equal to the size α of the test ϕ. The two tests have the same limiting power under the alternatives considered.

Similar results can be obtained for more general alternatives, for instance when the X_i are not identically distributed, provided only the central limit theorem can be applied.

5. An analysis of variance test. Let $\mathfrak{X}$ be a Euclidean space of np dimensions. Let $X = (X_1, \cdots, X_n)$ where $X_i = (X_{i1}, \cdots, X_{ip})$, $i = 1, \cdots, n$, are n independent random vectors of $p \geqq 2$ components, and let H be a hypothesis which implies that the distribution of each X_i is invariant under the $p!$ permutations of its components. Then the distribution of X is invariant under a group $\mathcal{G}$ of $M = (p!)^n$ permutations. For example, if in an agricultural experiment p treatments are randomly assigned to the p plots in each of n blocks, and X_{ij} is the yield of the plot in the ith block which has received the jth treatment, hypothesis H may be assumed to hold when there is no difference in the treatment effects.

Let the test $\phi(x)$ be defined by (1.1) with

$$t(x) = \frac{\sum_{j=1}^{p}\left(\sum_{i=1}^{n}(x_{ij} - x_{i.})\right)^2}{\sum_{i=1}^{n}(p-1)^{-1}\sum_{j=1}^{p}(x_{ij} - x_{i.})^2},$$

where $x_{i.} = p^{-1}\sum_{j=1}^{p} x_{ij}$. If the denominator vanishes, define $t(x) = p - 1$ (say). The denominator, which is invariant under permutations in $\mathcal{G}$, is so chosen that $Et(Gx) = p - 1$ for all x.

In the traditional analysis of variance one assumes that the X_{ij} are independent normal with common variance and means $EX_{ij} = b_i + t_j$. The equivalent of hypothesis H is that $t_1 = \cdots = t_p$. The usual F- (or z-) statistic for testing this hypothesis is an increasing function of $t(X)$.

A nonparametric test essentially equivalent to $\phi(x)$ was considered by Fisher [3] in the case $p = 2$, by Welch [14] and Pitman [11] in the general case.

Extending the customary alternative, suppose that

$$X_{ij} = Y_{ij} + b_i + t_j, \qquad i = 1, \cdots, n; \qquad j = 1, \cdots, p,$$

where the Y_{ij} are mutually independent and identically distributed,

$$EY_{ij} = 0, \qquad \operatorname{var} Y_{ij} = \sigma^2 > 0,$$

and the b_i and t_j are constants. It will be assumed that p is fixed and $n \to \infty$.

We can write

$$t(x) = \frac{\sum_{j=1}^{p} u_j(x)^2}{n^{-1} \sum_{i=1}^{n} (p-1)^{-1} \sum_{j=1}^{p} (x_{ij} - x_{i.})^2},$$

where

$$u_j(x) = n^{-\frac{1}{2}} \sum_{i=1}^{n} (x_{ij} - x_{i.}), \qquad j = 1, \cdots, p.$$

Since

$$X_{ij} - X_{i.} = Y_{ij} - Y_{i.} + t_j - \bar{t},$$

where $\bar{t} = p^{-1} \sum_1^p t_j$, has a distribution independent of i, the random variables

$$(p-1)^{-1} \sum_{j=1}^{p} (X_{ij} - X_{i.})^2, \qquad i = 1, \cdots, n,$$

are independent and identically distributed with mean $\sigma^2(1 + \delta^2)$, where

$$\delta^2 = \sigma^{-2}(p-1)^{-1} \sum_{k=1}^{p} (t_j - \bar{t})^2.$$

It follows that

$$n^{-1} \sum_{i=1}^{n} (p-1)^{-1} \sum_{j=1}^{p} (X_{ij} - X_{i.})^2 \to \sigma^2(1 + \delta^2) \text{ in probability.}$$

The expression on the left is invariant under the permutations in $\mathcal{G}$. Hence if we let

$$t'(x) = \sigma^{-2}(1 + \delta^2)^{-1} \sum_{j=1}^{p} u_j(x)^2,$$

then $(t(GX), t(G'X))$ has the same limiting distribution as $(t'(GX), t'(G'X))$.

We have

$$u_j(x) = \sum_{k=1}^{p} (\delta_{jk} - p^{-1}) v_k(x), \qquad v_k(x) = n^{-\frac{1}{2}} \sum_{i=1}^{n} (x_{ik} - b_i),$$

where δ_{jk} is Kronecker's delta. Let

$$V_i = v_j(GX), \qquad V'_j = v_j(G'X).$$

Then the random vector $n^{\frac{1}{2}}V = n^{\frac{1}{2}}(V_1 \cdots, V_p, V'_1, \cdots, V'_p)$ is the sum of n independent random vectors, each of which has the distribution of

$$Z^* = (Z_{R_1}, \cdots, Z_{R_p}, Z_{R'_1}, \cdots, Z_{R'_p}),$$

where $Z_1, \cdots, Z_p$ are independent, Z_j has the distribution of $Y_{ij} + t_j$, and $(R_1, \cdots, R_p)$ and $(R'_1, \cdots, R'_p)$ are two independent random vectors, independent of the Z_j, whose values are the $p!$ equally probable permutations of $(1, \cdots, p)$. By the central limit theorem for sums of identically distributed vectors, the limiting distribution of $V - EV$ is $2p$-variate normal with the covariance matrix of Z^*. We have

$$EZ^m_{R_j} = Ep^{-1}\sum_{k=1}^{p} Z^m_k, \quad j = 1, \cdots, p; \qquad m = 1, 2,$$

hence

$$EZ_{R_j} = \bar{t}, \qquad \operatorname{var} Z_{R_j} = \sigma^2(1 + (1 - p^{-1})\delta^2).$$

If $j \neq j'$,

$$EZ_{R_j}Z_{R_{j'}} = Ep^{-1}(p-1)^{-1}\sum\sum_{k\neq k'} Z_kZ_{k'} = p^{-1}(p-1)^{-1}\sum\sum_{k\neq k'} t_kt_{k'}$$
$$= p(p-1)^{-1}\bar{t}^2 - p^{-1}(p-1)^{-1}\sum t_k^2,$$

hence

$$\operatorname{cov}(Z_{R_j}, Z_{R_{j'}}) = -\sigma^2p^{-1}\delta^2, \qquad j \neq j'.$$

The $Z_{R'_{j'}}$ have the same distribution as the Z_{R_j}, and since $EZ_{R_j}Z_{R'_{j'}} = E(p^{-1}\sum Z_k)^2$ has the same value for all j, j', we have

$$\operatorname{cov}(Z_{R_j}, Z_{R'_{j'}}) = C \text{ (say)}, \qquad j, j' = 1, \cdots, p.$$

Hence

$$EV_j = EV'_j = n^{\frac{1}{2}}\bar{t},$$
$$\operatorname{var}(V_j) = \operatorname{var}(V'_j) = \sigma^2(1 + (1 - p^{-1})\delta^2),$$
$$\operatorname{cov}(V_j, V_{j'}) = \operatorname{cov}(V'_j, V'_{j'}) = -\sigma^2p^{-1}\delta^2, \qquad j \neq j',$$
$$\operatorname{cov}(V_j, V'_{j'}) = C.$$

Let $\| c_{ij} \|$ be an orthonormal $p \times p$ matrix with $c_{p1} = \cdots = c_{pp}$, and let

$$W_j = \sum_{k=1}^{p} c_{jk}V_k, \qquad W'_j = \sum_{k=1}^{p} c_{jk}V'_k.$$

Then

$$\sum_{j=1}^{p} u_j(GX)^2 = \sum_{j=1}^{p-1} W_j^2, \qquad \sum_{j=1}^{p} u_j(G'X)^2 = \sum_{j=1}^{p-1} W_j'^2.$$

For $j, j' \leqq p - 1$ we obtain

$$EW_j = EW'_j = 0, \qquad EW_jW'_{j'} = 0,$$
$$EW_jW_{j'} = EW'_jW'_{j'} = \delta_{jj'}\sigma^2(1 + \delta^2).$$

Hence the limiting distribution of $(W_1, \cdots, W_{p-1}, W'_1, \cdots, W'_{p-1})$ is that of $2p - 2$ independent normal variables, each with mean 0 and variance $\sigma^2(1 + \delta^2)$. It follows that the limiting distribution of $(t'(GX), t'(G'X))$, and hence of $(t(GX), t(G'X))$, is that of $(\chi^2_{p-1}, \chi'^2_{p-1})$, where χ^2_{p-1} and χ'^2_{p-1} are independent, each having the chi-square distribution with $p - 1$ degrees of freedom. By Theorem 3.2, $t^{(k)}(X) \to \lambda$ in probability, where $\Pr\{\chi^2_{p-1} > \lambda\} = \alpha$. The test is asymptotically as powerful as the conventional analysis of variance test of the same size α.

6. Two-sample test; tests of randomness. Let $\mathfrak{X}$ be the n-dimensional Euclidean space, and let H be the hypothesis that the n components of $X = (X_1, \cdots, X_n)$ are independent and identically distributed. Then the distribution of X is invariant under all $M = n!$ permutations of its components. Let $\phi(x)$ be the test (1.1) with

$$t(x) = \frac{\sum_1^n (a_i - \bar{a})x_i}{\left\{\sum_1^n (a_i - \bar{a})^2(n-1)^{-1}\sum_1^n (x_i - \bar{x})^2\right\}^{\frac{1}{2}}}, \tag{6.1}$$

where $a_1, \cdots, a_n$ are given numbers, not all equal, $\bar{a} = n^{-1}\sum_1^n a_i$, $\bar{x} = n^{-1}\sum_1^n x_i$. The numbers $a_i = a_{ni}$ may depend on n. If the denominator vanishes, that is, if $x_1 = \cdots = x_n$, define $t(x) = 0$. The denominator is invariant under all permutations, and is so chosen that $t(GX)$ has mean 0 and variance 1 (unless $x_1 = \cdots = x_n$).

If X has the probability density

$$(2\pi\sigma^2)^{-\frac{1}{2}n} \exp\left\{-(2\sigma^2)^{-1}\sum_1^n (x_i - a_i\xi - \eta)^2\right\} \tag{6.2}$$

and $T(x)$ denotes the standard t-statistic for testing $\xi = 0$, then

$$T(x) = (n-2)^{\frac{1}{2}}t(x)(n - 1 - t(x)^2)^{-\frac{1}{2}},$$

so that $T(x)$ is an increasing function of $t(x)$.

Lehmann and Stein [8] have shown that the test $\phi(x)$ is most powerful similar for testing H against the alternative (6.2) with $\xi > 0$, and that the test based on $|t(x)|$ is most stringent similar against (6.2) with $\xi \neq 0$. If

$$a_i = 1 \text{ for } i = 1, \cdots, m; \qquad a_i = 0 \text{ for } i = m + 1, \cdots, n, \tag{6.3}$$

then (6.2) is the probability density of two independent random samples from two normal distributions with common variance and means $\xi + \eta$ and η, and

the numerator of $t(x)$ is, apart from a constant factor, the difference of the two sample means. Essentially this test was proposed by Pitman [11].

We first consider a case where H is true.

THEOREM 6.1. *Let $t(x)$ be defined by (6.1), let $Z_1, \cdots, Z_n, \cdots$, be independent and identically distributed with $E \mid Z_1 \mid^3 < \infty$ and $\operatorname{var} Z_1 > 0$, and let $Z = Z^{(n)} = (Z_1, \cdots, Z_n)$. Then in order that for every real y*

$$F_n(y, Z) \to \Phi(y) \text{ in probability} \tag{6.4}$$

it is necessary and sufficient that either Z_1 be normally distributed or

$$\frac{\max_{1 \le i \le n} (a_i - \bar{a})^2}{\sum_{i=1}^{n} (a_i - \bar{a})^2} \to 0. \tag{6.5}$$

Results similar to Theorem 6.1 were obtained by Wald and Wolfowitz [13] and Noether [9], who gave sufficient conditions (stronger than those of Theorem 6.1) for $F_n(y, Z) \to \Phi(y)$ with probability one, which, of course, implies (6.4). An argument analogous to that employed by Wald, Wolfowitz, and Noether will be used in Sections 8–10 below to obtain alternative sufficient conditions for (6.4).

PROOF OF THEOREM 6.1. We may and shall assume that

$$\bar{a} = 0, \qquad \sum_1^n a_i^2 = 1, \qquad EZ_1 = 0, \qquad EZ_1^2 = 1. \tag{6.6}$$

Then

$$t(x) = \sum_1^n a_i x_i \left\{ (n-1)^{-1} \sum_1^n (x_i - \bar{x})^2 \right\}^{-\frac{1}{2}}.$$

Since $\bar{Z} \to 0$ and $n^{-1} \sum_1^n Z_i^2 \to 1$ in probability we have

$$(n-1)^{-1} \sum_1^n (Z_i - \bar{Z})^2 \to 1$$

in probability. Hence $(t(GZ), t(G'Z))$ has the same limiting distribution (if any) as $(u(GZ), u(G'Z))$, where

$$u(x) = \sum_1^n a_i x_i .$$

Let $gx = x_r = (x_{r_1}, \cdots, x_{r_n})$, where $r = (r_1, \cdots, r_n)$ is a permutation of $(1, \cdots, n)$. If R and R' are two independent random vectors, independent of Z, such that $\Pr\{R = r\} = \Pr\{R' = r\} = M^{-1}$ for all r, we can write $(GZ, G'Z) = (Z_R, Z_{R'})$. For any two permutations r, r' we have

$$u(Z_r) = \sum_1^n a_i Z_{r_i} = \sum_1^n a_{s_i} Z_i ,$$

$$u(Z_{r'}) = \sum_1^n a_i Z_{r'_i} = \sum_1^n a_{s'_i} Z_i ,$$

where s_i and s'_i are defined by

$$r_{s_i} = i, \qquad r'_{s'_i} = i, \qquad i = 1, \cdots, n.$$

First suppose that Z_1 is normally distributed. Then $u(Z_r)$ and $u(Z_{r'})$ have a bivariate normal distribution with means 0, variances 1, and correlation coefficient

$$\rho_{rr'} = Eu(Z_r)u(Z_{r'}) = \sum_1^n a_{s_i} a'_{s'_i}.$$

Thus

$$\Pr\{u(Z_r) \leqq y,\ u(Z_{r'}) \leqq y'\} = \Phi(y, y', \rho_{rr'}), \tag{6.7}$$

where

$$\Phi(y, y', \rho) = \int_{-\infty}^{y} \int_{-\infty}^{y'} (2\pi)^{-1}(1 - \rho^2)^{-\frac{1}{2}} \exp\left\{-\frac{u^2 - 2\rho uv + v^2}{2(1 - \rho^2)}\right\} du\, dv.$$

If both sides of (6.7) are summed over all r, r' and divided by M^2, we obtain

$$\Pr\{u(Z_R) \leqq y,\ u(Z_{R'}) \leqq y'\} = E\Phi(y, y', \rho_{RR'}). \tag{6.8}$$

The random variable $\rho_{RR'}$ has the same distribution as $\sum_1^n a_i a_{R_i}$, and we get $E\rho_{RR'}^2 = (n - 1)^{-1}$. Hence $\rho_{RR'} \to 0$ in probability.

Since $\Phi(y, y', \rho) \to \Phi(y)\Phi(y')$ as $\rho \to 0$, we have $\Phi(y, y', \rho_{RR'}) \to \Phi(y)\Phi(y')$ in probability. And since $\Phi(y, y', \rho)$ is a bounded function, this implies

$$E\Phi(y, y', \rho_{RR'}) \to \Phi(y)\Phi(y').$$

Hence $\Phi(y)\Phi(y')$ is the limiting distribution function of $(u(Z_R), u(Z_{R'}))$, and also that of $(t(GZ), t(G'Z))$. Relation (6.4) follows from Theorem 3.2.

Next suppose that (6.5) is satisfied. By assumption (6.6) this condition is equivalent to

$$\max_{1 \leqq i \leqq n} |a_i| \to 0. \tag{6.9}$$

If we let $Y_i = a_{s_i} Z_i$, $Y'_i = a'_{s'_i} Z_i$, the conditions of Theorem 3A are fulfilled, and we have $Y = u(Z_r)$, $Y' = u(Z_{r'})$, $\rho = \rho_{rr'}$, $\omega = \omega' = E|Z_1|^3 c_n$, where

$$c_n = \sum_1^n |a_i|^3.$$

Hence

$$|\Pr\{u(Z_r) \leqq y,\ u(Z_{r'}) \leqq y'\} - \Phi(y)\Phi(y')| \leqq g(\rho_{rr'}, c_n), \tag{6.10}$$

where the function $g(u, v)$ is independent of n and of the distribution of the Y_i, Y'_i (in particular, independent of r, r'), and $g(u, v) \to 0$ as $u \to 0$, $v \to 0$. Clearly $g(u, v)$ can be so defined that $g(u, v) \leqq 1$ for all u, v.

From (6.10) we obtain in a similar way as before

$$|\Pr\{u(Z_R) \leqq y,\ u(Z_{R'}) \leqq y'\} - \Phi(y)\Phi(y')| \leqq Eg(\rho_{RR'}, c_n). \tag{6.11}$$

Since $c_n \leqq \max | a_i | \sum_1^n a_i^2 = \max | a_i |$, condition (6.9) implies that $c_n \to 0$. Since $\rho_{RR'} \to 0$ in probability, $g(\rho_{RR'}, c_n) \to 0$ in probability; and since $g(u, v)$ is bounded, $Eg(\rho_{RR'}, c_n) \to 0$. Relation (6.4) now follows from (6.11) by Theorem 3.2.

Now suppose that Z_1 is not normal and (6.5) is not satisfied, the remaining assumptions of the theorem being fulfilled. Still assuming that the a_i satisfy (6.6), denote by A_n an a_j for which $| a_j | = \max (| a_1 |, \cdots, | a_n |)$. Then infinitely many $| A_n |$ are greater than a positive constant, and since the A_n are bounded, a subsequence $\{A_{n_m}\}$ of $\{A_n\}$ converges to a constant $A \neq 0$. We can write $u(Z) = u_n(Z) = V_n + W_n$, where V_n has the distribution of $A_n Z_1$ and is independent of W_n. As $m \to \infty$, V_{n_m} has the limiting distribution of AZ_1, which is not normal.

Suppose (6.4) were true. Then $t(Z_R)$, and hence $u_n(Z_R)$, would have the limiting distribution $\Phi(y)$. But Z_R has the same distribution as Z. It would follow that $u_{n_m}(Z)$ tends in distribution to a normal random variable which is the sum of two independent, nonnormal random variables. By a theorem of Cramér ([1], p. 52) this is impossible. The proof is complete.

In the sequel an extension by the author [7] of a theorem of Wald and Wolfowitz [13] will be required which, for purposes of reference, is stated below as Theorem 6A. For every positive integer n let $a = (a_1, \cdots, a_n)$, $b = (b_1, \cdots, b_n)$ be two vectors whose components a_i, b_i are real numbers and may depend on n. Suppose that the a_i are not all equal and the b_i are not all equal. Let the random vector $R = (R_1, \cdots, R_n)$ be defined as in the proof of Theorem 6.1, and let

$$F_n(y, a, b) = \Pr\left\{\frac{(n-1)^{\frac{1}{2}} \sum_1^n (a_i - \bar{a}) b_{R_i}}{\left[\sum_1^n (a_i - \bar{a})^2 \sum_1^n (b_i - \bar{b})^2\right]^{\frac{1}{2}}} \leqq y\right\},$$

where $\bar{a} = n^{-1} \sum_1^n a_i$, $\bar{b} = n^{-1} \sum_1^n b_i$.

THEOREM 6A. *A sufficient condition for*

$$F_n(y, a, b) \to \Phi(y) \tag{6.12}$$

as $n \to \infty$ is that

$$n^{\frac{1}{2}p-1} \frac{\sum_1^n (a_i - \bar{a})^p}{\left[\sum_1^n (a_i - \bar{a})^2\right]^{\frac{1}{2}p}} \frac{\sum_1^n (b_i - \bar{b})^p}{\left[\sum_1^n (b_i - \bar{b})^2\right]^{\frac{1}{2}p}} \to 0, \quad p = 3, 4, \cdots. \tag{6.13}$$

Condition (6.13) *is satisfied if*

$$n \frac{\max_{1 \leq i \leq n} (a_i - \bar{a})^2}{\sum_1^n (a_i - \bar{a})^2} \frac{\max_{1 \leq i \leq n} (b_i - \bar{b})^2}{\sum_1^n (b_i - \bar{b})^2} \to 0. \tag{6.14}$$

The next theorem is concerned with the behavior of $t^{(k)}(X)$ under an alternative which generalizes (6.2).

THEOREM 6.2. *Let $t(x)$ be defined by* (6.1), *and suppose that*

$$X_i = Z_i + d_i, \qquad i = 1, \cdots, n,$$

where $Z_1, \cdots, Z_n, \cdots$ are independent and identically distributed with

$$E\,|\,Z_1\,|^3 < \infty$$

and $\operatorname{var} Z_1 > 0$, *and $d_1, \cdots, d_n$ are constants (which may depend on n). Then*

$$t^{(k)}(X) \to \lambda \text{ in probability,} \tag{6.15}$$

where $\Phi(\lambda) = 1 - \alpha$, if

$$Z_1 \text{ is normal or } \frac{\max\limits_{1 \le i \le n} (a_i - \bar{a})^2}{\sum_1^n (a_i - \bar{a})^2} \to 0 \tag{6.16}$$

and

$$n^{\frac{1}{2}p - 1} \frac{\sum_1^n (a_i - \bar{a})^p}{\left[\sum_1^n (a_i - \bar{a})^2\right]^{\frac{1}{2}p}} \frac{\sum_1^n (d_i - \bar{d})^p}{\left[\sum_1^n (d_i - \bar{d})^2\right]^{\frac{1}{2}p}} \to 0, \qquad p = 3, 4, \cdots, \tag{6.17}$$

the latter condition being satisfied if

$$n \frac{\max\limits_{1 \le i \le n} (a_i - \bar{a})^2}{\sum_1^n (a_i - \bar{a})^2} \frac{\max\limits_{1 \le i \le n} (d_i - \bar{d})^2}{\sum_1^n (d_i - \bar{d})^2} \to 0. \tag{6.18}$$

Relation (6.15) *also holds if* (6.16) *is satisfied and*

$$n^{-1} \sum_1^n (d_i - \bar{d})^2 \to 0 \tag{6.19}$$

or if (6.17) *is satisfied and*

$$n^{-1} \sum_1^n (d_i - \bar{d})^2 \to \infty. \tag{6.20}$$

PROOF. We again make the simplifying assumptions (6.6). In addition we may set

$$\bar{d} = 0.$$

We then have

$$X_i - \bar{X} = Z_i - \bar{Z} + d_i,$$

$$n^{-1} \sum_1^n (X_i - \bar{X})^2 = n^{-1} \sum_1^n (Z_i - \bar{Z})^2 + D_n^2 + 2D_n s_n,$$

where

$$D_n = \left(n^{-1} \sum_1^n d_i^2\right)^{\frac{1}{2}}, \qquad s_n = \sum_1^n d_i Z_i \left(n \sum_1^n d_i^2\right)^{-\frac{1}{2}}.$$

We have $n^{-1} \sum_1^n (Z_i - \bar{Z})^2 \to 1$ in probability. Since $Es_n^2 = n^{-1}$, $s_n \to 0$ in probability. Also $0 \leqq 2D_n \leqq 1 + D_n^2$. Hence

$$\frac{n^{-1} \sum_1^n (X_i - \bar{X})^2}{1 + D_n^2} \to 1 \text{ in probability.}$$

Thus if we let

$$t'(x) = (1 + D_n^2)^{-\frac{1}{2}} \sum_1^n a_i x_i = (1 + D_n^2)^{-\frac{1}{2}} u(x),$$

then $(t(GX), t(G'X)) = (t(X_R), t(X_{R'}))$ has the same limiting distribution (if any) as $(t'(X_R), t'(X_{R'}))$.

Let

$$v(r) = \sum_1^n a_i d_{r_i} \left(n^{-1} \sum_1^n d_i^2\right)^{-\frac{1}{2}}.$$

Then

$$(6.21) \qquad t'(X_R) = \frac{u(Z_R) + D_n v(R)}{(1 + D_n^2)^{\frac{1}{2}}}, \qquad t'(X_{R'}) = \frac{u(Z_{R'}) + D_n v(R')}{(1 + D_n^2)^{\frac{1}{2}}}.$$

Suppose that conditions (6.16) and (6.17) are satisfied, and consider the joint distribution of $u(Z_R)$, $v(R)$, $u(Z_{R'})$, $v(R')$ as $n \to \infty$. It is seen from the proof of Theorem 6.1 that if (6.16) holds true and $p_n M^2$ denotes the number of pairs of permutations (r, r') for which $| \Pr \{u(Z_r) \leqq y, u(Z_{r'}) \leqq y'\} - \Phi(y)\Phi(y') |$ is less than a positive constant, then $p_n \to 1$ as $n \to \infty$. By the continuity theorem for the Fourier transform an analogous relation holds for the difference of the characteristic functions,

$$E \exp(itu(Z_r) + it'u(Z_{r'})) - \exp(-\tfrac{1}{2}t^2 - \tfrac{1}{2}t'^2).$$

Hence it follows that the characteristic function of $(v(R), v(R'), u(Z_R), u(Z_{R'}))$,

$$M^{-2} \sum_r \sum_{r'} \exp(i\tau v(r) + i\tau' v(r')) E \exp(itu(Z_r) + it'u(Z_{r'})),$$

differs arbitrarily little from

$$Ee^{i\tau v(R)} Ee^{i\tau' v(R')} e^{-\frac{1}{2}t^2 - \frac{1}{2}t'^2}$$

if n is sufficiently large. By Theorem 6A, condition (6.17) implies that $v(R)$ and $v(R')$ have the standard normal limiting distribution. Hence the limiting joint distribution of $v(R)$, $v(R')$, $u(Z_R)$, $u(Z_{R'})$ is that of four independent standard normal random variables. By (6.21) this implies that $(t'(X_R), t'(X_{R'}))$, and hence $(t(GX), t(G'X))$, has the limiting distribution function $\Phi(y)\Phi(y')$.

If (6.19) is satisfied, then $D_n \to 0$. Since $Ev(R)^2 = n(n-1)^{-1}$ is bounded, this implies that $D_n v(R) \to 0$ in probability, and $(t'(X_R), t'(X_{R'}))$ has the same limiting distribution as $(u(Z_R), u(Z_{R'}))$. When (6.16) holds, we can apply Theorem 6.1.

Similarly, if (6.20) is satisfied, $(t'(X_R), t'(X_{R'})$ has the limiting distribution of $(v(R), v(R'))$, which, under condition (6.17), is given by Theorem 6A. In every case the limiting distribution of $(t(GX), t(G'X))$ is $\Phi(y)\Phi(y')$, and relation (6.15) follows from Theorem 3.2. That condition (6.18) is sufficient for (6.17) is stated in Theorem 6A. This completes the proof.

If, in particular, X has the normal distribution (6.2), we have $d_i = a_i\xi + \eta$, and the conditions of Theorem 6.2 are fulfilled if either

$$n^{\frac{1}{2}(p-2)} \frac{\sum_1^n (a_i - \bar{a})^p}{\left[\sum_1^n (a_i - \bar{a})^2\right]^{\frac{1}{2}p}} \to 0, \qquad p = 3, 4, \cdots, \tag{6.22}$$

or

$$n^{\frac{1}{2}} \frac{\max_{1 \le i \le n} (a_i - \bar{a})^2}{\sum_1^n (a_i - \bar{a})^2} \to 0, \tag{6.23}$$

(which implies (6.22)), or

$$n^{-1} \sum_1^n (a_i - \bar{a})^2 \to 0. \tag{6.24}$$

In the two-sample case (6.3) the conditions (6.22) and (6.23) are both equivalent to

$$n(m')^{-2} = n^{-1}\left(\frac{m'}{n}\right)^{-2} \to 0,$$

where $m' = \min(m, n - m)$. Condition (6.24) is fulfilled if and only if

$$\frac{m'}{n} \to 0.$$

At least one of the two conditions is satisfied if m/n tends to some limit.

If the conditions of Theorem 6.2 up to and including (6.16) are satisfied, $t(X)$ is asymptotically normally distributed as $n \to \infty$. If the power of Student's (one-sided) test of size α tends to a limit, the power of ϕ tends to the same limit. Theorems 6.1 and 6.2 can be easily extended to the case where $Z_1, \cdots, Z_n$ have a common distribution which depends on n.

7. The two-sample test when one sample is small. It is of some interest to investigate what happens when the necessary and sufficient condition of The-

orem 6.1 is not satisfied. In the two-sample case, which will be discussed in this section, this occurs only if m or $n - m$ does not tend to infinity with n.

We first consider a somewhat more general situation. Let $\mathfrak{X}$ be the Euclidean n-dimensional space and $\mathcal{G}$ the group of all $M = n!$ permutations of the n coordinates of a point in $\mathfrak{X}$. Let the components $X_1, \cdots, X_n$ of X be independent. The function $t(x)$ can be arbitrary, subject only to the conditions to be stated.

First assume that $t(x) = u(x_1, \cdots, x_m)$ is a function of $x_1, \cdots, x_m$ only, where m is fixed as $n \to \infty$. The proportion of pairs of permutations r, r' for which the sets $(r_1, \cdots, r_m)$ and $(r_1', \cdots, r_m')$ have no elements in common tends to 1 as $n \to \infty$. Hence $t(X_r)$ and $t(X_{r'})$ are independent for a proportion of pairs r, r' which converges to 1. Suppose now that $X_1, \cdots, X_m$ have a common distribution and $X_{m+1}, \cdots, X_n$ have a common distribution. Then for a proportion of permutations r which tends to 1, $t(X_r)$ has the distribution function of $u(X_{m+1}, \cdots, X_{2m})$, which will be denoted by $F(y)$. It follows that $(t(X_R), t(X_{R'}))$ has the limiting distribution function $F(y)F(y')$. If the equation $F(y) = 1 - \alpha$ has a unique solution $y = \lambda$, $t^{(k)}(X) \to \lambda$ in probability by Theorem 3.2.

The same conclusions hold under the more general assumption that $t(x)$ is of the form $c(x)u(x_1, \cdots, x_m) + d(x)$, where $c(X_R) \to 1$ and $d(X_R) \to 0$ in probability, as follows from Theorem 3.3.

Now let $t(x)$ be defined by (6.1) with the a_i given by (6.3). Then

$$t(x) = \left\{\frac{m(n-m)}{n(n-1)} \sum_1^n (x_i - \bar{x})^2\right\}^{-\frac{1}{2}} \left(\sum_1^m x_i - m\bar{x}\right).$$

Suppose that m is fixed, and that the common distribution of $X_{m+1}, \cdots, X_n$ has mean μ and variance σ^2. Then

$$\bar{X} = n^{-1} \sum_1^n X_i \to \mu, \; n^{-1} \sum_1^n (X_i - \bar{X})^2 \to \sigma^2$$

in probability. Hence the preceding results can be applied with

$$u(x_1, \cdots, x_m) = m^{-\frac{1}{2}}\sigma^{-1} \sum_1^m (x_i - \mu).$$

Observe that the probability limit λ of $t^{(k)}(X)$ depends on the distribution of X_{m+1}. Now it follows from [8] that the two-sample test ϕ is most powerful similar for testing H not only against the normal alternative (6.2), (6.3), but also against any alternative with a density of the form

$$\prod_{i=1}^m f(x_i, \theta_1) \cdot \prod_{i=m+1}^n f(x_i, \theta_2), \tag{7.1}$$

where

$$f(y, \theta) = A(\theta)B(y)e^{\theta y}, \; \theta_1 > \theta_2 .$$

On the other hand, the most powerful test of size α for testing that $X_1, \cdots, X_n$ are independent with the common density $f(y, \theta)$, where

$$\theta = (m\theta_1 + (n - m)\theta_2)/n,$$

against (7.1) is of the form $\phi^*(x) = 1$ or 0 according as $m^{-\frac{1}{2}} \sum_1^m (x_i - \bar{x}) > c_n$ or $< c_n$, where $\sigma^{-1}c_n$ converges to the probability limit λ of $t^{(k)}(X)$. In other words, $t^{(k)}(X)$ always tends in probability to the "correct" value λ, so that the test ϕ is asymptotically as powerful as the most powerful parametric test for the case where the function $f(y, \theta)$ is known. This phenomenon is analogous to the relation between, say, the one-sided two-sample t-test for normal distributions with unknown variance σ^2 and the most powerful tests (corresponding to the different values of σ^2) when σ^2 is known.

8. An alternative approach. In the remaining part of the paper an alternative method of proving that $F_n(y, X)$ tends to $F(y)$ in probability will be considered. It is an extension of an argument used by Wald and Wolfowitz [13].

Suppose that the quantities $a_i = a_{ni}$, $b_i = b_{ni}$ in Theorem 6A are random variables which have a joint distribution for all i and all n, and suppose that one of the conditions (6.13), (6.14) is satisfied with probability 1. Then (6.12) holds with probability 1.

For example, let $X = (U_1, V_1, U_2, V_2, \cdots, U_n, V_n)$, where the pairs (U_i, V_i), $i = 1, \cdots, n$, are independent and identically distributed, and let H be the hypothesis that U_i and V_i are independent. When H is true, the distribution of X is invariant under the $M = (n!)^2$ permutations which permute $(U_1, \cdots, U_n)$ or $(V_1, \cdots, V_n)$. Let

$$t(x) = \frac{(n-1)^{\frac{1}{2}} \sum_1^n (u_i - \bar{u})(v_i - \bar{v})}{\left\{\sum_1^n (u_i - \bar{u})^2 \sum_1^n (v_i - \bar{v})^2\right\}^{\frac{1}{2}}}.$$

A test equivalent to the corresponding test $\phi(x)$ was considered by Pitman [11].

Since $t(x)$ is invariant under permutations of the pairs (u_i, v_i), the distribution of $t(Gx)$ is the same as the conditional distribution with $u_1, \cdots, u_n$ held in a fixed order and only the v_i permuted in all possible ways. Hence if in Theorem 6A we let $a_i = u_i$, $b_i = v_i$, then $F_n(y, a, b)$ is identical with $F_n(y, x)$. Suppose that U_1 and V_1 have finite moments of any order. Then the strong law of large numbers implies that condition (6.13) of Theorem 6A, with $a_i = U_i$, $b_i = V_i$, is satisfied with probability 1. Hence $F_n(y, X) \to \Phi(y)$ with probability 1, and a fortiori in probability. Theorem 3.1 can now be applied.

That condition (6.13) is satisfied with probability 1 can be shown under weaker assumptions (cf. Noether [9], [10]). Since, however, only weak convergence (in probability) of $F_n(y, X)$ is required for our purposes, a proof of strong convergence seems redundant. In fact, it will be shown in Sections 9 and 10 that if the conditions of Theorem 6A are satisfied as limits in probability, then the conclusion holds as a limit in probability.

9. Ordinary convergence and convergence in probability. Let $f_n(x_n)$, $g_n(x_n)$, $n = 1, 2, \cdots$, be two sequences of real-valued functions of elements x_n in a space $\mathfrak{X}_n$. Let X_n denote a random variable with values in $\mathfrak{X}_n$, $n = 1, 2, \cdots$.

THEOREM 9.1. *If $f_n(x_n) \to 0$ implies $g_n(x_n) \to 0$, then $f_n(X_n) \to 0$ in probability implies $g_n(X_n) \to 0$ in probability.*

PROOF. Suppose the theorem were false. Then there exist two sequences of functions $\{f_n\}$, $\{g_n\}$ such that $f_n(x_n) \to 0$ implies $g_n(x_n) \to 0$, and a sequence of random variables $\{X_n\}$ such that $f_n(X_n) \to 0$ in probability but, for some $\delta > 0$ and some $\epsilon > 0$, $\Pr\{|g_n(X_n)| > \delta\} > \epsilon$ for infinitely many n. Let m be an arbitrary positive integer. Consider the events

$$A_n = \{|g_n(X_n)| > \delta\}, \qquad B_n^{(m)} = \{|f_n(X_n)| < m^{-1}\}.$$

We have $\Pr\{A_n\} > \epsilon$ for infinitely many n, and there exists a number N_m such that $\Pr\{B_n^{(m)}\} > 1 - \frac{1}{2}\epsilon$ for $n > N_m$. If $A_n \cdot B_n^{(m)}$ denotes the joint occurrence of A_n and $B_n^{(m)}$,

$$\Pr\{A_n \cdot B_n^{(m)}\} \geqq \Pr\{A_n\} + \Pr\{B_n^{(m)}\} - 1 > \epsilon + 1 - \tfrac{1}{2}\epsilon - 1 > 0$$

for infinitely many n.

Hence for every positive integer m there exists a sequence $\{x_n^{(m)}\}$, $x_n^{(m)} \in \mathfrak{X}_n$, such that $|f_n(x_n^{(m)})| < m^{-1}$ for $n > N_m$ and $|g_n(x_n^{(m)})| > \delta$ for infinitely many n. For every $m = 1, 2, \cdots$ there exists an integer $K_m \geqq N_{m+1}$ such that

$$|g_{K_m}(x_{K_m}^{(m)})| > \delta$$

and $K_1 < K_2 < \cdots$. Let $K_0 = 0$,

$$x_n' = x_n^{(m)} \text{ for } n = K_{m-1} + 1, \cdots, K_m; \qquad m = 1, 2, \cdots.$$

Then $|f_n(x_n')| < m^{-1}$ for $n > K_m$, hence $f_n(x_n') \to 0$, and $|g_n(x_n')| > \delta$ for infinitely many n. But this contradicts the assumption.

Let, in particular, x_n be the vector (a, b) of Theorem 6A, f_n the left-hand side of (6.14) and $g_n = F_n(y, a, b) - \Phi(y)$. Then Theorem 9.1 shows that if a and/or b are replaced by random vectors, the fulfilment of (6.14) in probability implies that (6.12) holds in probability. Theorem 9.1 does not suffice to draw the same conclusion if the infinitely many relations (6.13) are satisfied in probability. That the conclusion is permissible will be shown in Section 10.

We conclude this section by stating, without proof, conditions which imply the fulfilment of (6.14) in probability. It can be shown that

$$n^{1-(2/h)} \frac{\max_{1 \leq i \leq n} (X_i - \bar{X})^2}{\sum_1^n (X_i - \bar{X})^2} \to 0 \text{ in probability} \tag{9.1}$$

if $X_1, \cdots, X_n, \cdots$ are independent, identically distributed, $E|X_1|^h < \infty$ for some $h \geqq 2$. Relation (9.1) with $h = 2$ also holds if $X_1, \cdots, X_n$ are independent with common mean and finite second moments and satisfy the Lindeberg condition of the central limit theorem. More generally, (9.1) holds if $EX_i = d_i$, the $X_i' = X_i - d_i$ satisfy one of the previously stated conditions and

$$n^{1-(2/h)} \frac{\max_{1 \leq i \leq n} (d_i - \bar{d})^2}{\sum_1^n (d_i - \bar{d})^2} \to 0.$$

Hence one can obtain alternative sufficient conditions for $t_n^{(k)}(X) \to \lambda$ in probability in the examples of Sections 6 and 8. Thus in the case of Section 8 it is sufficient that U_i and V_i have finite moments of order 4.

10. The second limit theorem for random distributions. A generalization by Fréchet and Shohat [5] of Markov's so-called second limit theorem of probability theory states that if the distribution function $F(y)$ is uniquely determined by its moments and $\{F_n(y)\}$ is a sequence of distribution functions whose moments converge to the corresponding moments of $F(y)$, then $F_n(y) \to F(y)$ at every point of continuity of $F(y)$. An extension of this theorem to the case where the $F_n(y)$ are random distribution functions and ordinary convergence is replaced by convergence in probability was given by M. N. Ghosh [6] under certain additional assumptions concerning $F(y)$ and its moments. The following theorem shows that the extension holds with no restrictions.

THEOREM 10.1. *Let $F(y)$ be a distribution function on the real line which is uniquely determined by its moments*

$$\mu_k = \int_{-\infty}^{\infty} y^k \, dF(y), \qquad k = 1, 2, \cdots.$$

Let $\{F_n(y)\}$, $n = 1, 2, \cdots$, be a sequence of random distribution functions with moments μ_{nk}, and suppose that

$$\mu_{nk} \to \mu_k \text{ in probability as } n \to \infty, \qquad k = 1, 2, \cdots.$$

Then

$$F_n(y) \to F(y) \text{ in probability}$$

at every point of continuity of $F(y)$.

The proof is based on the following lemma.

LEMMA 10.1. *Let $F(y)$ be a distribution function which is uniquely determined by its moments μ_k, $k = 1, 2, \cdots$. Then for every y' at which $F(y)$ is continuous and for every $\epsilon > 0$ there exist a positive integer $m = m(y', \epsilon)$ and a positive number $\delta = \delta(y', \epsilon)$ such that for every distribution function $G(y)$ whose moments ν_k satisfy the inequalities*

$$| \nu_k - \mu_k | < \delta, \qquad k = 1, \cdots, m,$$

we have

$$| G(y') - F(y') | < \epsilon.$$

PROOF.[2] Assume the lemma to be false. Then for some y' at which $F(y)$ is continuous and for some $\epsilon > 0$ there do not exist positive numbers m, δ for

[2] The author is indebted to H. Robbins for the proof of Lemma 10.1.

which the conclusion of the lemma holds. Hence for every positive integer m there exists a distribution function $G_m(y)$ with moments ν_{mk} such that

$$|\nu_{mk} - \mu_k| < m^{-1}, \qquad k = 1, \cdots, m,$$

and

$$|G_m(y') - F(y')| \geqq \epsilon.$$

But $\{G_m(y)\}$, $m = 1, 2, \cdots$, is a sequence of distribution functions whose moments, ν_{mk}, converge to μ_k for all $k = 1, 2, \cdots$. By the aforementioned theorem of Fréchet and Shohat, $G_m(y') \to F(y')$, which leads to a contradiction.

Proof of Theorem 10.1. Let y' be a point of continuity of $F(y)$. Given $\epsilon > 0$, let $m = m(y', \epsilon)$, $\delta = \delta(y', \epsilon)$ be defined as in Lemma 10.1. Given $\eta > 0$, choose N so that

$$\Pr \{|\mu_{nk} - \mu_k| < \delta, k = 1, \cdots, m\} > 1 - \eta \qquad \text{for } n > N.$$

It follows from Lemma 10.1 that $|F_n(y') - F(y')| < \epsilon$ with probability $> 1 - \eta$ for $n > N$. The proof is complete.

It will now be shown that if the relations (6.13) are satisfied as limits in probability, (6.12) holds in probability. It can be seen from the proof of Theorem 6A in [7] that if (6.13) holds for $p = 3, 4, \cdots k$, then the moments up to order k of the distribution $F_n(y, a, b)$ converge to the corresponding moments of $\Phi(y)$. By Theorem 9.1 this implies that if (6.13) holds in probability for every $p = 3, 4, \cdots$, then every moment of $F_n(y, a, b)$ converges in probability to the corresponding moment of $\Phi(y)$. By Theorem 10.1, $F_n(y, a, b) \to \Phi(y)$ in probability.

Relations (6.13) can be shown to hold in probability under conditions which are slightly weaker than those indicated at the end of Section 9, though the gain does not seem to be considerable.

REFERENCES

[1] H. Cramér, *Random Variables and Probability Distributions, Cambridge Tracts in Mathematics, No. 36*, Cambridge University Press, 1937.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] R. A. Fisher, *The Design of Experiments*, 5th ed., Hafner Publishing Co., New York, 1949.

[4] M. Fréchet, *Généralités sur les probabilités. Variables aléatoires.* Gauthier-Villars, Paris, 1937.

[5] M. Fréchet and J. Shohat, "A proof of the generalized second limit theorem in the theory of probability," *Trans. Am. Math. Soc.*, Vol. 33 (1931), p. 533–543.

[6] M. N. Ghosh, "Convergence of random distribution functions," *Bull. Calcutta Math. Soc.*, Vol. 42 (1950), p. 217–226.

[7] W. Hoeffding, "A combinatorial central limit theorem," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 558–566.

[8] E. L. Lehmann and C. Stein, "On the theory of some non-parametric hypotheses," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 28–45.

[9] G. E. Noether, "On a theorem by Wald and Wolfowitz," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 455–458.

[10] G. E. Noether, "Asymptotic properties of the Wald-Wolfowitz test of randomness," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 231–246.

[11] E. J. G. Pitman, "Significance tests which may be applied to samples from any population," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 4 (1937), pp. 119–130.

"Significance tests which may be applied to samples from any population. II. The correlation coefficient test," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 4 (1937), pp. 225–232.

"Significance tests which may be applied to samples from any population. III. The analysis of variance test," *Biometrika*, Vol. 29 (1938), pp. 322–335.

[12] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill Book Co., 1937.

[13] A. Wald and J. Wolfowitz, "Statistical tests based on permutations of the observations," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 358–372.

[14] B. L. Welch, "On the z-test in randomized blocks and Latin squares," *Biometrika*, Vol. 29 (1937), pp. 21–52.

ASYMPTOTIC THEORY OF CERTAIN "GOODNESS OF FIT" CRITERIA BASED ON STOCHASTIC PROCESSES

T. W. ANDERSON AND D. A. DARLING[1]

Columbia University and University of Michigan

1. Summary. The statistical problem treated is that of testing the hypothesis that n independent, identically distributed random variables have a specified continuous distribution function $F(x)$. If $F_n(x)$ is the empirical cumulative distribution function and $\psi(t)$ is some nonnegative weight function $(0 \leqq t \leqq 1)$, we consider $n^{\frac{1}{2}} \sup_{-\infty<x<\infty} \{| F(x) - F_n(x) | \psi^{\frac{1}{2}}[F(x)]\}$ and $n\int_{-\infty}^{\infty} [F(x) - F_n(x)]^2 \psi[F(x)]\, dF(x)$. A general method for calculating the limiting distributions of these criteria is developed by reducing them to corresponding problems in stochastic processes, which in turn lead to more or less classical eigenvalue and boundary value problems for special classes of differential equations. For certain weight functions including $\psi = 1$ and $\psi = 1/[t(1 - t)]$ we give explicit limiting distributions. A table of the asymptotic distribution of the von Mises ω^2 criterion is given.

2. Introduction. One method of testing the hypothesis that n observations have been drawn from a population with specified distribution function $F(x)$ is to compare the empirical histogram based on dividing the real line into intervals with the hypothetical histogram by means of the χ^2 tests. A test which does not involve a subjective grouping of the data consists of comparing the empirical cumulative distribution function with the hypothetical distribution function. Let $F_n(x)$ be the empirical distribution function based on n observations; that is, $F_n(x) = k/n$ if k observations are $\leqq x$ for $k = 0, 1, \cdots, n$. We wish to consider a convenient measure of the discrepancy or "distance" between two distribution functions. (For a more detailed discussion cf. Wald and Wolfowitz [21].) In accordance with the usual notions of a metric in function space, we treat the following measures:

$$n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2\psi[F(x)]dF = W_n^2, \tag{2.1}$$

$$\sup_{-\infty<x<\infty} \sqrt{n}\, | F_n(x) - F(x) | \sqrt{\psi[F(x)]} = K_n, \tag{2.2}$$

where $\psi(t)$ $(\geqq 0)$ is some preassigned weight function.

If a measure W_n^2 is adopted, the hypothesis is rejected for those samples for which $W_n^2 > z_1$, say, and if a measure K_n is adopted, the hypothesis is rejected when $K_n > z_2$, say. The numbers z_1 and z_2 are to be chosen so that when the hypothesis is true the probability of rejection is some specified number (for

[1] This work was done mainly at the Rand Corporation.

example, .01 or .05). The main purpose of this paper is to give methods for finding the asymptotic distributions of W_n^2 and K_n, and, hence, approximate values of the significance points, z_1 and z_2. We assume that the hypothetical distribution is continuous.

The fundamental ideas for tests of this nature are due to Kolmogorov [11], Smirnov [17], Cramér [2], and von Mises [19], and for large n certain tests have been developed by them. The present paper treats these tests in somewhat more detail, the analysis being greatly expedited by reducing the problems to straightforward considerations in the theory of continuous Gaussian stochastic processes. This reduction was developed by Doob [6], and used by him to give a simplified proof of Kolmogorov's fundamental result.

The principal innovation in this paper is the incorporation of a weight function to allow more flexibility in the tests. Although we are able to make explicit calculations for only a few simple types of weight functions, the principal mathematical problems are reduced to classical problems in the theory of differential equations.

The function $\psi(t)$, $0 \leqq t \leqq 1$, is to be chosen by the statistician so as to weight the deviations according to the importance attached to various portions of the distribution function. This choice depends on the power against the alternative distributions considered most important. The selection of $\psi(t) \equiv 1$ yields $n\omega^2$, the criterion of von Mises, for W_n^2, and the criterion of Kolmogorov for K_n. For W_n^2 to exist for all samples except a set with probability zero, it is necessary and sufficient that the following integrals exist:

$$\int_0^{u_1} u^2\psi(u)du \tag{2.3}$$

for every $u_1 (0 < u_1 < 1)$,

$$\int_{u_2}^1 (1-u)^2\psi(u)du \tag{2.4}$$

for every u_2 $(0 < u_2 < 1)$.

Given the data $x_1, x_2, \cdots, x_n$ arranged in increasing magnitude (with probability one there are no equalities between any two of them, since the distribution is assumed continuous), we obtain for practical computations the simpler variants of (2.1) and (2.2),

$$W_n^2 = 2\sum_{j=1}^n \left\{\phi_2[F(x_j)] - \frac{2j-1}{2n}\phi_1[F(x_j)]\right\} + n\int_0^1 (1-t)^2\psi(t)dt, \tag{2.5}$$

$$K_n = \frac{1}{\sqrt{n}} \max_{j=1,\cdots,n} \{\sqrt{\psi[F(x_j)]}\max[nF(x_j) - (j-1), j - nF(x_j)]\}, \tag{2.6}$$

where

$$\phi_1(t) = \int_0^t \psi(s)ds, \qquad \phi_2(t) = \int_0^t s\psi(s)ds. \tag{2.7}$$

For (2.5) to hold the integrals $\phi_1(t)$, $\phi_2(t)$ must exist; for (2.6) to hold it is necessary and sufficient that

$$\frac{1}{\psi(t)}\left|\frac{d}{dt}[t(1-t)\psi(t)]\right| \leqq 1 \tag{2.8}$$

if $\psi(t)$ is differentiable (substituting the difference quotient in (2.8) if $\psi(t)$ is not differentiable).

3. Reduction to a continuous stochastic process. Since $F(x)$ is assumed continuous, we can make the transformation $u = F(x)$. Then the observations are $u_i = F(x_i)$ $(i = 1, 2, \cdots, n)$, and under the null hypothesis these can be considered as drawn from the uniform distribution between 0 and 1. Let $G_n(u)$ be the empirical distribution derived from $u_1, \cdots, u_n$. Then W_n^2 and K_n are, respectively,

$$W_n^2 = n\int_0^1 [G(u) - u]^2\psi(u)\,du, \tag{3.1}$$

$$K_n = \sup_{0\leqq u\leqq 1} \sqrt{n}\,|\,G_n(u) - u\,|\,\sqrt{\psi(u)}. \tag{3.2}$$

For every $0 \leqq u \leqq 1$, $Y_n(u) = \sqrt{n}[G_n(u) - u)]$ is a random variable and the set of these random variables may be considered a stochastic process with parameter u. Putting

$$A_n(z) = \Pr\left\{\int_0^1 Y_n^2(u)\psi(u)du \leqq z\right\} = \Pr\{W_n^2 \leqq z\}, \tag{3.3}$$

$$B_n(z) = \Pr\{\sup_{0\leqq u\leqq 1} |\,Y_n(u)\,|\,\sqrt{\psi(u)} \leqq z\} = \Pr\{K_n \leqq z\}, \tag{3.4}$$

we wish to calculate $A(z) = \lim A_n(z)$, $n \to \infty$, and $B(z) = \lim B_n(z)$, $n \to \infty$, if these limits exist.

For fixed $u_1, u_2, \cdots, u_k$ the joint distribution of $Y_n(u_1), Y_n(u_2), \cdots, Y_n(u_k)$ approaches a k-variate normal distribution as $n \to \infty$. Thus the asymptotic process is Gaussian (normal) and is specified by its mean and covariance functions. For finite n we have

$$\begin{aligned} E(Y_n(u)) &= 0, \\ E(Y_n(u)Y_n(v)) &= \min(u, v) - uv. \end{aligned} \tag{3.5}$$

The limiting process is a Gaussian process $y(u)$, $0 \leqq u \leqq 1$, for which

$$\begin{aligned} E(y(u)) &= 0, \\ E(y(u)y(v)) &= \min(u, v) - uv, \end{aligned} \tag{3.6}$$

such that the probability is 1 that $y(u)$ is continuous [6]. Putting

$$a(z) = \Pr\left\{\int_0^1 y^2(u)\psi(u)\,du \leqq z\right\}, \tag{3.7}$$

$$b(z) = \Pr\{\sup_{0 \leq u \leq 1} |y(u)| \sqrt{\psi(u)} \leqq z\}, \tag{3.8}$$

we wish to conclude that $A(z) = a(z)$ and $B(z) = b(z)$. Having established these equalities we shall be in a position to use certain representation theorems for stochastic processes to great advantage.

In [4] Donsker has given the following theorem: *Let R be the space of real, single-valued functions $g(t)$ which are continuous except for at most a finite number of finite jumps, and let C be the space of continuous functions. Let $F(g)$ be a functional defined on R and continuous in the uniform topology, i.e.,* $\sup_{0 \leq t \leq 1} |g_n(t) - g_0(t)| \to 0$, $n \to \infty$, *implies* $|F(g_n) - F(g_0)| \to 0$, $n \to \infty$, $g_n \in R$, $g_0 \in C$, *except for a set of $g_0(t)$ with 0 probability according to the probability associated with $y(t)$. Then*

$$\lim_{n \to \infty} \Pr\{F[Y_n(t)] \leqq z\} = \Pr\{F[y(t)] \leqq z\}. \tag{3.9}$$

It follows from this result that if $\psi(u)$ is bounded $A(z) = a(z)$ and $B(z) = b(z)$.

To handle more general weight functions for the case of integrals we want to extend this result. We shall assume that $\psi(u)$ is continuous in any interval $0 < u_1 \leqq u \leqq u_2 < 1$. Secondly we assume that

$$\int_0^{u_1} \psi(t) t \log\log\frac{1}{t}\, dt, \qquad \int_{u_1}^1 \psi(t)(1 - t)\log\log\frac{1}{1 - t}\, dt \tag{3.10}$$

exist for every u_1 $(0 < u_1 < 1)$. It is shown in Section 5 that

$$(1 + t)y(t/(1 + t)) = X(t)$$

is the Wiener process which has the property ([12] p. 242 and p. 247)

$$\Pr\left\{\text{there exists a } t_0 \text{ such that } X^2(t) \leqq 2t\log\log\frac{1}{t} \text{ for } 0 < t < t_0\right\} = 1. \tag{3.11}$$

This implies that

$$\Pr\left\{\begin{aligned}&\text{there exists a } u_0 \text{ such that}\\ &\qquad y^2(u) \leq 2u(1 - u)\log\log\frac{1 - u}{u} \text{ for } 0 < u < u_0\end{aligned}\right\} = 1. \tag{3.12}$$

Thus with probability 1 $\psi(t)y^2(t)$ is majorized by $k\psi(t)t\log\log(1/t)$ for $k \geqq 2(1 - u_0)$. Thus if the first integral in (3.10) exists

$$\int_0^{u_1} y^2(t)\psi(t)\, dt \tag{3.13}$$

exists with probability 1 (taking the principal value when the integral is improper). A similar argument holds for the existence of

$$\int_{u_1}^1 y^2(t)\psi(t)\, dt. \tag{3.14}$$

Thus $\int_0^1 y^2(t)\psi(t)\,dt$ exists with probability 1. This defines a functional continuous in the uniform topology. Hence from Donsker's theorem $A(z) = a(z)$.

4. The limiting distribution of the integral criterion. In this section we show how to find $a(z)$ in terms of the solution of a certain differential equation and give two examples of this method. The statistic W_n^2 is essentially that introduced by Cramér [2]; in the case of $\psi(t) \equiv 1$, it is n times the ω^2 criterion studied by von Mises [19] and Smirnov [17].

The method we use is analogous to the technique of Kac and Siegert [10]. We shall sketch briefly the extension of their results.

By Mercer's theorem a symmetric continuous correlation function $k(s, t)$, $0 \leq s, t \leq 1$, which is square integrable (in one variable and in both variables), can be expressed as

$$k(s, t) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} f_j(s) f_j(t), \tag{4.1}$$

where λ_i is an eigenvalue and $f_i(t)$ is the corresponding normalized eigenfunction of the integral equation

$$\lambda \int_0^1 k(s, t) f(s)\,ds = f(t), \tag{4.2}$$

and

$$\int_0^1 f_i(t) f_j(t)\,dt = \delta_{ij}, \tag{4.3}$$

the Kronecker delta. In most cases $k(0, 0) = k(1, 1) = 0$; hence $f_i(0) = f_i(1) = 0$. Since $k(s, t)$ is positive definite, $\lambda_i > 0$. The series (4.1) converges absolutely and uniformly in the unit square.

Let $X_1, X_2, \cdots$ be independently, normally distributed with means zero and variances 1. If $k(t, t) < \infty$, then we can define

$$z(t) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} X_j f_j(t); \tag{4.4}$$

the series converges in the mean and with probability one for each t. Then $z(t)$ is a Gaussian process with $Ez(t) = 0$ and $Ez(s)z(t) = k(s, t)$. Thus $z(t)$ gives the same stochastic process as $\sqrt{\psi(t)}\, y(t)$ when $k(s, t) = \sqrt{\psi(s)}\sqrt{\psi(t)}$ $[\min(s, t) - st]$. From this it follows that with probability 1

$$\begin{aligned} W^2 &= \int_0^1 \psi(t) y^2(t)\,dt = \int_0^1 z^2(t)\,dt = \int_0^1 \left[\sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} f_j(t) X_j\right]^2 dt \\ &= \sum_{j=1}^{\infty} \frac{1}{\lambda_j} X_j^2. \end{aligned} \tag{4.5}$$

See [10] for details of this proof. Thus

$$
\begin{aligned}
E[e^{iuW^2}] &= E\left[\exp\left(iu \sum_{j=1}^{\infty} X_j^2/\lambda_j\right)\right] \\
&= \prod_{j=1}^{\infty} E[\exp iuX_j^2/\lambda_j] \qquad (4.6)\\
&= \prod_{j=1}^{\infty} (1 - 2iu/\lambda_j)^{-\frac{1}{2}}.
\end{aligned}
$$

The infinite product converges absolutely and uniformly for all real u, and in general $1/\lambda_n = O(1/n^2)$.

We desire a more general result, however, because one weight function we treat leads to a kernel that is not continuous at (0, 0) and (1, 1). We use the following theorem of Hammerstein [9]: *Let $k(s, t)$ be continuous in the unit square except possibly at the corners of the square; let $\partial k(s, t)/\partial s$ be continuous in the interior of both triangles in which the square is divided by the line between* (0, 0) *and* (1, 1), *and let the partial derivative be bounded in the domain $\epsilon \leqq s \leqq 1 - \epsilon$ and $0 \leqq t \leqq 1$ for each $\epsilon(> 0)$. Then the series on the right of* (4.1) *converges uniformly to $k(s, t)$ in every domain in the interior of the unit square.*

Since $k(s, t) = \sqrt{\psi(s)}\sqrt{\psi(t)}\,[\min(s, t) - st]$, the condition is that $\psi(t)$ be continuous for $0 < t < 1$ and

$$\sqrt{\frac{\psi(t)}{\psi(s)}}\; t[\tfrac{1}{2}(1 - s)\psi'(s) - \psi(s)] \tag{4.7}$$

be continuous for $0 \leqq t \leqq s \leqq 1 - \epsilon$ and

$$\sqrt{\frac{\psi(t)}{\psi(s)}}\; (1 - t)[\tfrac{1}{2}s\psi'(s) + \psi(s)] \tag{4.8}$$

be continuous for $\epsilon \leqq s \leqq t \leqq 1$.

In this case (4.4) converges in the mean and with probability one for every $t(\epsilon \leqq t \leqq 1 - \epsilon)$, and $z(t)$ is the same process as $x(t)$ in this interval. If $\int_0^1 k(t, t)\, dt < \infty$, $\sum_{j=1}^{\infty} 1/\lambda_j < \infty$ (by Bessel's inequality) and $\sum_{j=1}^{\infty} X_j^2/\lambda_j$ converges with probability one. Further, with probability one, $\sum_{j=1}^{\infty} X_j f_j(t)/\sqrt{\lambda_j}$ converges in the mean (integral with respect to t) and it converges to $z(t)$. Thus we have with probability one

$$\int_0^1 z^2(t)\, dt = \sum_{j=1}^{\infty} X_j^2/\lambda_j, \tag{4.9}$$

$$\int_\epsilon^{1-\epsilon} x^2(t)\, dt = \int_\epsilon^{1-\epsilon} z^2(t)\, dt = \int_\epsilon^{1-\epsilon} \left[\sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} X_j f_j(t)\right]^2 dt. \tag{4.10}$$

For ϵ small enough

$$E\left[\int_0^\epsilon x^2(t)\, dt + \int_{1-\epsilon}^{1} x^2(t)\, dt\right] = \int_0^\epsilon k(t, t)\, dt + \int_{1-\epsilon}^1 k(t, t)\, dt < \delta \tag{4.11}$$

for any $\delta > 0$. Thus the distribution of $W^2 = \int_0^1 x^2(t)\,dt$ is the limiting distribution of $\int_\epsilon^{1-\epsilon} x^2(t)\,dt$. With a similar argument for the integral of $z^2(t)$ we argue that the distribution of W^2 is the distribution of $\sum_{j=1}^\infty X_j^2/\lambda_j$ with characteristic function (4.6).

THEOREM 4.1. *If*

$$k(s, t) = \sqrt{\psi(s)}\,\sqrt{\psi(t)}\,[\min(s, t) - st] \tag{4.12}$$

is continuous or if $k(s, t)$ *is continuous except at* (0, 0) *and* (1, 1) *with* $\partial k(s, t)/\partial s$ *continuous for* $0 < s, t < 1$, $s \neq t$, *and bounded in* $\epsilon \leqq s \leqq 1 - \epsilon$, $0 \leqq t \leqq 1$ *for every* ϵ (>0) *then the characteristic function of* W^2 *is given by* (4.6), *where* $\{\lambda_j\}$ *are the eigenvalues of* $k(s, t)$ *defined by* (4.2).

In our case the integral equation is

$$f(t) = \lambda \int_0^1 [\min(t, s) - ts]\,\sqrt{\psi(t)}\,\sqrt{\psi(s)}\,f(s)\,ds. \tag{4.13}$$

It can be shown that if $f(t)$ satisfies (4.13) for some λ, then $h(t) = f(t)\psi^{-\frac{1}{2}}(t)$ satisfies

$$h''(t) + \lambda\psi(t)h(t) = 0 \tag{4.14}$$

for that λ (see [8], Sections 604 and 605) and $h(0) = h(1) = 0$ when $k(0, 0) = k(1, 1) = 0$. Let $h(t, \lambda)$ be the solution of (4.14) for which

$$h(0, \lambda) = 0, \qquad \left.\frac{\partial h(t, \lambda)}{\partial t}\right|_{t=0} = 1. \tag{4.15}$$

If $\psi(t)$ is continuous $(0 \leqq t \leqq 1)$, such a solution exists and $h(t, \lambda)$ is continuous in $t(0 \leqq t \leqq 1)$. Since $h(1, \lambda) = 0$ for λ an eigenvalue of (4.13), the roots of $h(1, \lambda) = 0$ are the roots of the Fredholm determinant $D(\lambda)$ associated with $k(s, t)$. It can be shown that

$$D(\lambda) = \frac{h(1, \lambda)}{h(1, 0)} = \prod_{i=1}^\infty \left(1 - \frac{\lambda}{\lambda_i}\right). \tag{4.16}$$

The characteristic function (4.6) is

$$\frac{1}{\sqrt{D(2it)}}. \tag{4.17}$$

The square root is taken so as to make (4.17) real and positive when the characteristic function is real and positive. The details of this proof are given in [8], Section 605.

THEOREM 4.2. *Let* $\psi(t)$ *be continuous for* $0 \leqq t \leqq 1$. *Then the equation* (4.14) *has a unique solution* $h(t, \lambda)$ *for every* $\lambda > 0$ *satisfying* (4.15). *Then the characteristic function of* W^2 *is*

$$\sqrt{\frac{h(1, 0)}{h(1, 2it)}}. \tag{4.18}$$

Thus we have reduced the problem of finding the characteristic function of W^2 to finding the general solution of a differential equation.

The semi-invariants κ_n of W^2 are given quite easily (when they exist) through the eigenvalues. Since

$$\phi(t) = \prod_{j=1}^{\infty} (1 - 2it/\lambda_j)^{-1}, \tag{4.19}$$

the coefficient of $(it)^n/n!$ in the power series expansion of $\log \phi(t)$ is

$$\kappa_n = 2^{n-1}(n-1)! \sum_{j=1}^{\infty} \left(\frac{1}{\lambda_j}\right)^n, \qquad n = 1, 2, \cdots. \tag{4.20}$$

Hence we obtain for the mean and variance, for instance,

$$\begin{aligned} \kappa_1 &= \mu = \sum \frac{1}{\lambda_j}, \\ \kappa_2 &= \sigma^2 = 2 \sum \left(\frac{1}{\lambda_j}\right)^2. \end{aligned} \tag{4.21}$$

Even without knowing the eigenvalues, the moments can be calculated in terms of the iterates of the kernel $k(s, t)$. Putting $k_1(s, t) = k(s, t) = (\min(s, t) - st)\sqrt{\psi(s)\psi(t)}$, $k_{n+1}(s, t) = \int_0^1 k_n(s, u)k(u, t)\, du$, we have by means of the bilinear expansion

$$k_n(s, t) = \sum \lambda_j^{-n} f_j(s) f_j(t). \tag{4.22}$$

Hence,

$$\kappa_n = 2^{n-1}(n-1)! \int_0^1 k_n(s, s)\, ds \tag{4.23}$$

and, in particular,

$$\begin{aligned} \mu &= \int_0^1 k(s, s)\, ds = \int_0^1 s(1-s)\psi(s)\, ds, \\ \sigma^2 &= 2\int_0^1 \int_0^1 k^2(s, t)\, ds\, dt = 4 \int_0^1 (1-s)^2 \psi(s) \int_0^s t^2 \psi(t)\, dt\, ds. \end{aligned} \tag{4.24}$$

We now present two applications of this method.

Example 1. Let $\psi(t) \equiv 1$; then $W_n^2 = n\omega^2$. The differental equation $h''(t) + \lambda h(t) = 0$ has a solution

$$h(t, \lambda) = \frac{1}{\sqrt{\lambda}} \sin \sqrt{\lambda} t \tag{4.25}$$

satisfying (4.15). Taking $h(1, 0)$ as $\lim_{\lambda \to 0} h(1, \lambda) = 1$, we find that $1/\sqrt{D(2it)}$ is

$$\phi_1(t) = Ee^{itw^2} = \sqrt{\frac{\sqrt{2it}}{\sin \sqrt{2it}}} = \sqrt{\frac{\sqrt{-2it}}{\sinh \sqrt{-2it}}}. \tag{4.26}$$

This expression was given by Smirnov [15] and later by von Mises [20] using entirely different methods. A formal method for finding the distribution (by inverting the Fourier transform) was given later by Smirnov [16], but his expression is not amenable to numerical calculation. The following procedure expresses $a_1(z) = \Pr\{W^2 \leqq z\}$ in terms of tabulated functions.

It appears convenient to work with the Laplace transform. We have

$$\xi(t) = \phi_1(it) = E(e^{-tW^2}) = \sqrt{\frac{\sqrt{2t}}{\sinh \sqrt{2t}}}. \tag{4.27}$$

Using integration by parts, we obtain

$$\int_0^\infty e^{-zt}\, a_1(z)\, dz = \frac{1}{t}\,\xi(t) \tag{4.28}$$

for the cdf $a_1(z)$. We wish to invert this Laplace transform. Now

$$\frac{1}{t}\,\xi(t) = \left(\frac{2}{t}\right)^{3/4} e^{-\frac{1}{2}\sqrt{2t}}\,(1 - e^{-2\sqrt{2t}})^{-\frac{1}{2}}. \tag{4.29}$$

We suppose in the sequel that the real part of t, $R(t) > 0$ and apply the binomial expansion to the last expression; thus

$$\frac{\xi(t)}{t} = \left(\frac{2}{t}\right)^{3/4} \sum_{j=0}^{\infty} (-1)^j \binom{-\frac{1}{2}}{j} e^{-(2j+\frac{1}{2})\sqrt{2t}}, \tag{4.30}$$

where $\binom{-\frac{1}{2}}{j} = (-1)^j\Gamma(j + \frac{1}{2})/[\Gamma(\frac{1}{2})j!]$. It may be readily verified that the complex inversion formula can be used termwise here since the abscissa of convergence of $\xi(t)/t$ is $R(t) = 0$, and the above series converges absolutely and uniformly in the half plane $R(t) \geqq \beta > 0$.

Since

$$\begin{aligned} e^{-A\sqrt{t}} &= \int_0^\infty e^{-st} \frac{A}{2\sqrt{\pi}\, s^{3/2}} e^{-A^2/(4s)}\, ds, \\ \frac{1}{t^{3/4}} &= \int_0^\infty e^{-st} \frac{ds}{\Gamma(3/4)s^{1/4}}, \end{aligned} \tag{4.31}$$

we have

$$\frac{e^{-A\sqrt{t}}}{t^{3/4}} = \int_0^\infty e^{-zt}\,\phi(z)\, dz, \tag{4.32}$$

where

$$\phi(z) = \frac{A}{2\sqrt{\pi}\,\Gamma(3/4)} \int_0^z \frac{e^{-A^2/(4x)}}{x^{3/2}(z - x)^{1/4}}\, dx \tag{4.33}$$

by virtue of the convolution property of the Laplace transform. In this integral we change variables, putting $x = u \operatorname{sech}^2 \theta$ to give

$$
\begin{aligned}
\phi(z) &= \frac{A}{\sqrt{\pi}\,\Gamma(3/4)z^{3/4}} \int_0^\infty e^{-(A^2/(4z))\cosh^2\theta}(\cosh\theta\sinh\theta)^{\frac{1}{2}}\,d\theta \\
&= \frac{Ae^{-A^2/(8z)}}{2^{3/2}\sqrt{\pi}\,\Gamma(3/4)z^{3/4}} \int_0^\infty e^{-(A^2/(8z))\cosh\theta}(\sinh\theta)^{\frac{1}{2}}\,d\theta \qquad (4.34)\\
&= \frac{e^{-A^2/(8z)}}{\sqrt{2}\,\pi}\sqrt{\frac{A}{z}}\,K_{\frac{1}{4}}\left(\frac{A^2}{8z}\right),
\end{aligned}
$$

where $K_{\frac{1}{4}}(x)$ is the standard Bessel function [22].

Having inverted the typical term, we finally obtain by summing

$$
\begin{aligned}
a_1(z) = \frac{1}{\pi\sqrt{z}} \sum_{j=0}^\infty (-1)^j \binom{-\frac{1}{2}}{j} \qquad (4.35)\\
\cdot(4j+1)^{\frac{1}{2}} e^{-(4j+1)^2/(16z)} K_{\frac{1}{4}}((4j+1)^2/(16z)).
\end{aligned}
$$

The convergence of this series is very rapid. If $a_1(z) = \sum_{j=0}^\infty u_j(z)$, we find that $u_{j+1}(z)/u_j(z) < k_j e^{-(4j+1)/(2z)}$ (using the fact that $K_{\frac{1}{4}}(t)$ is a decreasing function of t), where $k_0 < 1.12$, $k_1 < 1.007$, $k_2 < 1.002$, $k_j < 1.0007$ for $j \geqq 3$. Since $K_{\frac{1}{4}}(t)$ is positive, $u_j(z) > 0$. Using a crude geometric series bound for $R_4(z) = \sum_{j=4}^\infty u_j(z)$, we can show that for $z \leqq 2$, $R_4(z) < .0002$. Moreover, for $z \leqq 2$, $R_4(z) < u_3(z) < u_2(z) < u_1(z)$. In computation, therefore, one need only take as many terms in the series as are different from 0 in the number of decimal places carried. We give below a table of z for equal increments (.01) of $a_1(z)$ with the 5%, 1% and .1% significance points. The calculations have been carried to 6 figures before rounding off. The authors are indebted to Mr. Jack Laderman of Columbia University and the Numerical Analysis Department of the Rand Corporation for their assistance in preparing the table.

The semi-invariants of this distribution are easily obtained since the eigenvalues are $\lambda_j = 1/(\pi^2 j^2)$. Thus

$$
\begin{aligned}
\kappa_n &= \frac{2^{n-1}(n-1)!}{\pi^{2n}} \sum_{j=1}^\infty \frac{1}{j^{2n}} \qquad (4.36)\\
&= 2^{3n-2}\frac{(n-1)!}{(2n)!} B_n,
\end{aligned}
$$

where B_n are the Bernoulli numbers: $B_1 = 1/6$, $B_2 = 1/30$, etc.

Example 2. $\psi(t) = 1/[t(1-t)]$. Since the variance of $Y_n(t) = \sqrt{n}\,[G_n(t) - t]$ is $t(1-t)$, an interesting weight function for $Y_n^2(t)$ is the reciprocal of this variance.[2] In a certain sense, this function assigns to each point of the distribution

[2] This suggestion was first made by L. J. Savage.

TABLE 1

Limiting Distribution of $n\omega^2$

$$a_1(z) = \lim_{n\to\infty} \Pr\{n\omega^2 \le z\}$$

z	$a_1(z)$	z	$a_1(z)$	z	$a_1(z)$
.02480	.01	.08562	.34	.17159	.67
.02878	.02	.08744	.35	.17568	.68
.03177	.03	.08928	.36	.17992	.69
.03430	.04	.09115	.37	.18433	.70
.03656	.05	.09306	.38	.18892	.71
.03865	.06	.09499	.39	.19371	.72
.04061	.07	.09696	.40	.19870	.73
.04247	.08	.09896	.41	.20392	.74
.04427	.09	.10100	.42	.20939	.75
.04601	.10	.10308	.43	.21512	.76
.04772	.11	.10520	.44	.22114	.77
.04939	.12	.10736	.45	.22748	.78
.05103	.13	.10956	.46	.23417	.79
.05265	.14	.11182	.47	.24124	.80
.05426	.15	.11412	.48	.24874	.81
.05586	.16	.11647	.49	.25670	.82
.05746	.17	.11888	.50	.26520	.83
.05904	.18	.12134	.51	.27429	.84
.06063	.19	.12387	.52	.28406	.85
.06222	.20	.12646	.53	.29460	.86
.06381	.21	.12911	.54	.30603	.87
.06541	.22	.13183	.55	.31849	.88
.06702	.23	.13463	.56	.33217	.89
.06863	.24	.13751	.57	.34730	.90
.07025	.25	.14046	.58	.36421	.91
.07189	.26	.14350	.59	.38331	.92
.07354	.27	.14663	.60	.40520	.93
.07521	.28	.14986	.61	.43077	.94
.07690	.29	.15319	.62	.46136	.95
.07860	.30	.15663	.63	.49929	.96
.08032	.31	.16018	.64	.54885	.97
.08206	.32	.16385	.65	.61981	.98
.08383	.33	.16765	.66	.74346	.99
				1.16786	.999

$F(x)$ equal weights. A statistician may prefer to use this weight function when he feels that $\psi(t) \equiv 1$ does not give enough weight to the tails of the distribution.

In this example

$$(4.37)\qquad \begin{aligned} k(t, s) &= \sqrt{\frac{t(1-s)}{(1-t)s}}, \qquad t \leqq s, \\ &= \sqrt{\frac{(1-t)s}{t(1-s)}}, \qquad t \geqq s, \end{aligned}$$

is not continuous at $(t, s) = (0, 0)$ or $(1, 1)$; hence we need the extended result of Theorem 4.1 to justify our procedure. It is known that the Ferrer associated Legendre polynomials $f_i(t) = P_i^1(t) = t(1 - t)P_i(2t - 1)$ satisfy the integral equation with $\lambda_i = 1/[i(i + 1)]$ (see [23], p. 324). Thus the characteristic function of W^2 is

$$(4.38)\qquad \begin{aligned} \phi_2(t) &= \prod_{j=1}^{\infty}\left(1 - \frac{2it}{j(j+1)}\right)^{-\frac{1}{2}} \\ &= \sqrt{\frac{-2\pi it}{\cos\left(\frac{\pi}{2}\sqrt{1+8it}\right)}}. \end{aligned}$$

An analysis similar to that used in Example 1 shows that the cdf, $a_2(z)$, can be expressed as

$$\begin{aligned} a_2(z) &= \Pr\{W^2 \leqq z\} \\ &= \sqrt{\frac{\pi}{2}}\frac{1}{z}\sum_{j=0}^{\infty}\binom{-\frac{1}{2}}{j}(4j+1)\int_0^1 e^{(rz)/8-((4j+1)^2\pi^2)/(8rz)}\frac{dr}{r^{3/2}(1-r)^{1/2}} \\ &= \frac{\sqrt{2\pi}}{z}\sum_{j=0}^{\infty}\binom{-\frac{1}{2}}{j}(4j+1)e^{-((4j+1)^2\pi^2)/(8z)}\int_0^{\infty} e^{z/(8(w^2+1))-((4j+1)^2\pi^2w^2)/(8z)}\,dw. \end{aligned}$$

5. Theory of deviations. The second test criterion led to the calculation of

$$B_n(z) = \Pr\{\sup_{0 \leqq u \leqq 1} \sqrt{n}\mid G_n(u) - u \mid \sqrt{\psi(u)} \leqq z\}.$$

In order to handle the limiting distribution we consider the functional

$$(5.1)\qquad K = \sup_{0 \leqq u \leqq 1} \mid y(u) \mid \sqrt{\psi(u)}.$$

It follows from the theorem of Donsker [4] that for $\psi(u)$ bounded we have

$$\lim_{n\to\infty} B_n(z) = \Pr\{K \leqq z\},$$

and the problem is reduced to that of calculating the distribution of (5.1). This is the elegant idea of Doob [6], who treated the case $\psi \equiv 1$.

This is known as an "absorption probability" problem on account of its very suggestive analogy with a simple diffusion model. It is clear that the event that $\{-z(\psi(u))^{-\frac{1}{2}} \leqq y(u) \leqq z(\psi(u))^{-\frac{1}{2}}, 0 \leqq u \leqq 1\}$ is equivalent to the event $\{K \leqq z\}$; thus the probability $b(z)$ is, very crudely speaking, the "proportion" of all those

paths $y(u)$ of the diffusing particle which do not get "absorbed into" (i.e., intersect) the "barriers" $y = \pm z(\psi(u))^{-\frac{1}{2}}$.

It is convenient to make a transformation due to Doob [6] which renders the analysis simpler. If we put

$$X(t) = (1 + t)y\left(\frac{t}{1 + t}\right),$$

it is easy to verify that $X(t)$ is the Wiener-Einstein process; that is, $X(t)$ is Gaussian, $X(0) = 0$, $E(X(t)) = 0$, $E(X(t)X(s)) = \min(s, t)$. Then

$$b(z) = \Pr\{| X(t) | \leqq \xi(t), 0 \leqq t \leqq \infty\},$$

where

$$\xi(t) = \frac{z(1 + t)}{\sqrt{\psi\left(\frac{t}{1 + t}\right)}}. \tag{5.2}$$

Thus we have the absorption probability problem for the free particle with barriers $x = \pm\xi(t)$ for $t \geqq 0$.

The method of solution is to treat the corresponding diffusion problem as a boundary value problem with the diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2}\frac{\partial^2 f}{\partial x^2} \tag{5.3}$$

associated with the region $t \geqq 0$, $| x | \leqq \xi(t)$. In line with the preceding analogy $f(t, x)$ will be the "density" of paths $X(u)$ which for $0 \leqq u \leqq t$ have not been "absorbed" and for which $X(t) = x$; hence

$$\int_{|x|<\xi(t)} f(t, x)\, dx$$

will give the probability of nonabsorption up to time t. It is the limit of this expression for $t \to \infty$ which will yield $b(z)$. For a more detailed discussion of these points see Lévy [12], pp. 78 et seq.

We need the following existence and uniqueness theorem:

THEOREM 5.1. *Given that $\xi(t)$ of (5.2) has a bounded derivative for $t_0 \leqq t \leqq t_1$, there exists a unique function $p(t_0, y; t, x)$ such that for any continuous function $g(y)$, $| y | < \xi(t_0)$, the function*

$$f(t, x) = \int_{|y|<\xi(t_0)} g(y)p(t_0, y; t, x)\, dy \tag{5.4}$$

has the following properties:

(1) *$f(t, x)$ satisfies (5.3) in the domain $t_0 < t < t_1$, $| x | < \xi(t)$,*

(2) $$\lim_{x\to\pm\xi(t)} f(t, x) = 0, \qquad t_1 > t > t_0,$$

$$(3) \qquad \lim_{\substack{t \to t_0 \\ x \to \eta}} f(t, x) = g(\eta) \qquad | \eta | < \xi(t_0).$$

The proof of this theorem is contained quite explicitly in the fundamental paper of Fortet [7] (especially ch. V), who considers in great detail the general problem of absorption probabilities. Fortet treats only the case of one absorbing barrier, but his results are easily extended to the above case of two barriers. The differential $p(t_0, y; t, x)\,dx$ can be interpreted, to terms of order $(dx)^2$, as the probability that if the diffusing particle starts at (t_0, y) it will not have been absorbed in the barriers $\pm\xi(t)$ during the interval (t_0, t), and will lie between x and $x + dx$ at time t.

We have not stated the best theorem possible. If $\xi(t)$ is merely continuous the absorption probability density $f(t, x)$ exists. For the existence of a solution to (5.3) satisfying (2) and (3) of Theorem 5.1 it is sufficient to require that $\xi(t)$ satisfy a Lipschitz condition associated with the law of the iterated logarithm. Finally we remark in passing that unless $f(t, x)$ is of the form (5.4) (the so called "normal" solution of Fortet) its uniqueness is not assured (cf. Doetsch [3]).

If in the theorem $\xi(t)$ has a bounded derivative for $t \geqq 0$ then we plainly have

$$(5.5) \qquad b(z) = \lim_{t \to \infty} \int_{-\xi(t)}^{\xi(t)} p(0, 0; t, x)\,dx,$$

but if $\xi(t)$ does not have a bounded derivative for $t \geqq 0$, (5.5) can no longer be employed to determine $b(z)$. However, if there are a finite number of intervals in each of which $\xi(t)$ has a bounded derivative and between which $\xi(t)$ has a simple jump discontinuity it is easy to modify the above result; in fact over some of the intervals $\xi(t)$ may be infinite. A piecewise determination can be made and the solution can be continued to beyond the last discontinuity, and then (5.5) can be used. Suppose the points of discontinuity of $\xi(t)$ are $0 < t_1 < t_2 < \cdots < t_n$ and suppose $\xi(t)$ is, say, left continuous. In the region $(0, t_1)$ we have the solution $g_0(t, x) = p_0(0, 0; t, x)$ by the above theorem. Now if $\xi(t_1) < \xi(t_1 + 0)$ we define $g_1^*(t_1, x)$ by

$$g_1^*(t_1, x) = \begin{cases} g_0(t_1, x), & | x | \leqq \xi(t_1), \\ 0, & \xi(t_1) \leqq | x | \leqq \xi(t_1 + 0), \end{cases}$$

and if $\xi(t_1) > \xi(t_1 + 0)$ we define $g_1^*(t_1, x) = g_0(t_1, x)$, $| x | \leqq \xi(t_1 + 0)$. Then $g_1^*(t_1, x)$ is continuous in $| x | < \xi(t_1 + 0)$ and we have for $t_1 < t < t_2$ a function $g_1(t, x)$ defined by Theorem 5.1;

$$g_1(t, x) = \int_{|y| < \xi(t_1 + 0)} g_1^*(t_1, y) p_1(t_1, y; t, x)\,dy.$$

In the same way we can define a function $g_2^*(t_2, x)$ which will yield a function

$g_2(t, x)$ for $t_2 < t < t_3$. This process will ultimately yield a unique function $g_n(t, x)$ for $t > t_n$. Finally

$$b(z) = \lim_{t\to\infty} \int_{-\xi(t)}^{\xi(t)} g_n(t, x)\, dx. \tag{5.6}$$

It is clear that if $\xi(t) = \infty$ in some of the intervals the successive determination of the functions $g_k(t, x)$ may still be carried forward. This would correspond to an absence of the absorbing barrier over the interval.

Using the relation (5.2) and the above remarks we have the following theorem for the weight function $\psi(u)$:

THEOREM 5.2. *Suppose there is a finite sequence* $0 = u_0 < u_1 < u_2 \cdots < u_n < u_{n+1} = 1$ *such that in the interval* $(u_k, u_{k+1}]$ $\psi(t)$ *is either* (1) *identically zero or* (2) *is bounded away from zero and has a bounded derivative. Then there exists a unique sequence of functions* $\{p_k(t_k, y; t, x)\}$ *such that for* t *in the interval* $((u_k/(1 - u_k) = t_k < t < t_{k+1} = u_{k+1}/(1 - u_{k+1}))$ *the conclusions of Theorem* 5.1 *hold for the functions* $p_k(t_k, y; t, x)$, $k = 0, 1, \cdots, n$, $\xi(t)$ *being defined by* (5.2).

From this theorem we can generate a set of functions $g_k(t, x)$, $t_k < t < t_{k+1}$, $k = 0, 1, \cdots, n$, and another set $g_k^*(t_k, x)$, $k = 1, 2, \cdots, n$, as before. $g_{k+1}^*(t_{k+1}, x)$ agrees with $g_k(t_{k+1}, x)$ over the set of x for which the latter is defined; that is, $|x| < \xi(t_{k+1})$, and is zero for other values; namely, $\xi(t_{k+1} + 0) > |x| > \xi(t_{k+1})$ if $\xi(t)$ has a positive jump at t_{k+1}. Putting

$$g_0(t, x) = p_0(0, 0; t, x), \qquad t \leq t_1,$$

$$g_k(t, x) = \int_{|y| < \xi(t_k + 0)} g_k^*(t_k, y) p_k(t_k, y; t, x)\, dy, \quad t_k < t < t_{k+1}, \quad k = 1, 2, \ldots n,$$

we finally have (5.6) for $b(z)$.

In a formal way the problem is thus solved, but the analytical difficulties of getting an explicit solution may be prohibitive. If $\xi(t)$ consists of a set of linear arcs (which implies that $\sqrt{\psi(u)}$ is of the form $(\alpha u + \beta)^{-1}$ in a piecewise way) then $b(z)$ can be determined by quadratures (see, for example, Goursat [8], ch. 29, Ex. 3). We make an application of this remark below.

It is clear that if $\psi(u)$ becomes infinite for some $0 < u < 1$ then $b(z) = 0$ for every $z > 0$. But since $y(0) = y(1) = 0$ it is possible that $\psi(u)$ may become infinite for $u = 0$ or 1 and still yield a nondegenerate $b(z)$. But in this case it is necessary that $\psi(u)$ not dominate $[2u(1 - u) \log \log 1/(u(1 - u))]^{-1}$ for u near 0 or 1.

We shall consider several examples.

Example 1. Let $\psi(u)$ be a constant over a set of intervals,

$$\psi(u) = q_k \geq 0, \qquad u_k < u \leq u_{k+1}, \qquad u_0 = 0, \qquad u_{n+1} = 1, \qquad k = 0, 1, \cdots n.$$

By choosing enough intervals, an arbitrary weight function can be approximated, in a manner of speaking.

It follows that the problem will be essentially solved if we can determine the

functions $p_k(t_k, y; t, x)$ of Theorem 5.2. In this case the function $\xi(t)$ becomes, by (5.2),

$$\xi(t) = \frac{z}{\sqrt{q_k}}(1 + t), \qquad \frac{u_k}{1 - u_k} < t \leqq \frac{u_{k+1}}{1 - u_{k+1}},$$

and we must find the solution to equation (5.3) which satisfies the conditions (2) and (3) of Theorem 5.1.

As before we put $t_k = u_k/(1 - u_k)$, and it follows by a classical procedure of superposing an infinite system of sources and sinks along the line $t = t_k$ that we may get the Green's solution. In fact, let us put a source at $t = t_k$, $x = y_j$, of strength s_j, where

$$y_j = 2j \frac{z}{\sqrt{q_k}}(t_k + 1) + (-1)^j y,$$

$$s_j = (-1)^j \exp\left\{-2\frac{z^2}{q_k}(t_k + 1)j^2 - 2\frac{z}{\sqrt{q_k}} yj(-1)^j\right\}$$

for $j = 0, \pm 1, \pm 2, \cdots$. Then for $t_k < t \leqq t_{k+1}$ and $|y| < (z/\sqrt{q_k})(1 + t_k)$ we obtain

$$p_k(t_k, y; t, x) = \sum_{j=-\infty}^{\infty} \frac{s_j}{\sqrt{2\pi(t - t_k)}} e^{-\frac{1}{2}(x-y_j)^2/(t-t_k)}, \tag{5.7}$$

which may be directly verified by substitution to be a solution. It has been tacitly assumed that $q_k > 0$; if $q_k = 0$ we obtain only the term corresponding to $j = 0$ in the above solution, namely, the fundamental solution

$$p(t_k, y; t, x) = \frac{1}{\sqrt{2\pi(t - t_k)}} e^{-\frac{1}{2}(x-y)^2/2(t-t_k)}.$$

Now on putting

$$r_k = \min\left\{\frac{z}{\sqrt{q_k}}(1 + t_k), \frac{z}{\sqrt{q_{k-1}}}(1 + t_k)\right\}, \qquad k = 1, 2, \cdots, n,$$

and using the method outlined above, we obtain

$$g_n(t, x) = \int_{-r_n}^{r_n} \cdots \int_{-r_2}^{r_2} \int_{-r_1}^{r_1} p_0(0, 0; t_1, x_1) p_1(t_1, x_1; t_2, x_2) \cdots p_n(t_n x_n; t, x) \, dx_1 \, dx_2 \cdots dx_n$$

for $p_k(t_k, x_k; t_{k+1}, x_{k+1})$ as in (5.7), and finally as an "explicit" solution,

$$b_1(z) = \lim_{t \to \infty} \int_{|x| < \frac{z}{\sqrt{q_n}}(1+t)} g_n(t, x) \, dx.$$

The resulting function $b_1(z)$ is a multiply infinite sum of integrals of an n-variate Gaussian distribution over an n-dimensional rectangle.

We consider now the following special case of the above result

$$\psi(u) = \begin{cases} 1, & 0 \leq a < u \leq b \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus the test of the hypothesis is confined to detecting discrepancies over only a central portion of the interval [0, 1]. Using the preceding notation we have $n = 2$ and

$$u_0 = 0, \quad t_0 = 0, \quad q_0 = 0,$$

$$u_1 = a, \quad t_1 = \frac{a}{1-a}, \quad q_1 = 1,$$

$$u_2 = b, \quad t_2 = \frac{b}{1-b}, \quad q_2 = 0,$$

and hence

$$p(0, 0; t_1, x_1) = \frac{e^{-(x_1^2)/(2t_1)}}{\sqrt{2\pi t_1}},$$

$$p_1(t_1, x_1; t_2, x_2) = \sum_{j=-\infty}^{\infty} \frac{s_j}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{1}{2}(x_2 - y_j)^2/(t_2 - t_1)},$$

$$\text{(5.8)} \quad \begin{cases} y_j = 2jz(t_1 + 1) + (-1)^j x_1, \\ s_j = (-1)^j \exp\{-2z^2(t_1 + 1)j^2 - 2zx(-1)^j\}, \end{cases}$$

$$p_2(t_2, x_2; t, x) = \frac{e^{-\frac{1}{2}(x - x_2)^2/(t - t_2)}}{\sqrt{2\pi(t - t_2)}}.$$

Thus, putting $b_1(z) = P(a, b, z)$,

$$P(a, b, z) = \lim_{t \to \infty} \int_{-\infty}^{\infty} \int_{-z(1+t_2)}^{z(1+t_2)} \int_{-z(1+t_1)}^{z(1+t_1)} p(0, 0; t_1, x_1)\, p_1(t_1, x_1; t_2, x_2) p_2(t_2, x_2; t, x)\, dx_1\, dx_2\, dx$$

$$= \int_{-z(1+t_2)}^{z(1+t_2)} \int_{-z(1+t_1)}^{z(1+t_1)} p(0, 0; t_1, x_1) p_1(t_1, x_1; t_2, x_2) dx_1\, dx_2$$

$$= \sum_{j=-\infty}^{\infty} \frac{s_j}{\sqrt{2\pi t_1(t_2 - t_1)}} \int_{-z(1+t_1)}^{z(1+t_1)} \int_{-z(1+t_2)}^{z(1+t_2)} \exp\left(-\frac{x_1^2}{2t_1} - \frac{(x_2 - y_j)^2}{2(t_2 - t_1)}\right) dx_2\, dx_1$$

for s_j and y_j as in (5.8).

The double integral is seen to be over a bivariate normal distribution, and if we let $n(x_1, x_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ be the normal bivariate density in x_1, x_2 with

means μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation ρ we obtain by rewriting the above integral

$$P(a, b, z) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2z^2j^2} \int_{-z(1+t_1)}^{z(1+t_1)} \int_{-z(1+t_2)}^{z(1+t_2)} n(x_1, x_2, \mu_1^{(j)}, \mu_2^{(j)}, \sigma_1^2, \sigma_2^2, \rho_j) dx_2 dx_1,$$

where

$$\mu_1^{(j)} = -2zj(-1)^j t_1, \quad \mu_2^{(j)} = 2zj, \quad \sigma_1^2 = t, \quad \sigma_2^2 = t_2, \quad \rho_j = (-1)^j \sqrt{\frac{t_1}{t_2}}.$$

A somewhat simpler way of writing this result is as follows. Let $M(u, v, \xi, \eta, \rho)$ be the volume under the normal bivariate surface with means zero and variances 1 and correlation ρ which is above the rectangle with vertices

$$x = u \pm \xi,$$
$$y = v \pm \eta.$$

Then, remembering that $t_1 = a/(1 - a)$, $t_2 = b/(1 - b)$ and $M(u, v, \xi, \eta, \rho) = M(-u, v, \xi, \eta, -\rho)$, we obtain after a simple transformation of the above integral

$$(5.9) \qquad \begin{aligned} P(a, b, z) &= \sum_{j=-\infty}^{\infty} (-1) e^{-2z^2j^2} \\ &\cdot M\left(2jz\sqrt{\frac{a}{1-a}}, 2jz\sqrt{\frac{1-b}{b}}, \frac{z}{\sqrt{a(1-a)}}, \frac{z}{\sqrt{b(1-b)}}, -\sqrt{\frac{a(1-b)}{b(1-a)}}\right). \end{aligned}$$

There are tables available in which the function M is tabulated; see also Pólya [14]. Also, if either $a = 0$ or $b = 1$ then $\rho = 0$ and the function can be calculated with the ordinary univariate Gaussian tables. Putting $a = 0$, $b = 1$ simultaneously we obtain Kolmogorov's result

$$P(0, 1, z) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2z^2},$$

which has been tabulated [18]. In the general case the convergence is very rapid and good results can be obtained by using a few central terms (in (5.9) the terms corresponding to $\pm j$ are clearly equal).

The formula (5.9) is in disagreement with a recent announcement (without proof) of Maniya [13]. Maniya's note appeared subsequent to a restricted paper by the authors.

By using the general formula above it is possible to get, for example, a weight function to test discrepancies over only the tails of the distribution, etc.

Example 2. We next investigate

$$\psi(u) = \begin{cases} \dfrac{1}{u(1-u)}, & 0 < a < u \leqq b < 1, \\ 0, & \text{otherwise,} \end{cases}$$

which is the weight function considered before with the W^2 test. By an earlier remark we must have $a > 0$ and $b < 1$, else absorption is certain and $b(z)$ is degenerate. The transformation (5.2) yields

$$b_2(z) = \Pr\left\{ |\, X(t)\,| < z\sqrt{t},\ \frac{a}{1-a} < t < \frac{b}{1-b} \right\},$$

where $X(t)$ is the Wiener-Einstein process. Here it appears convenient to make another transformation. Let $u(t)$ be the Uhlenbeck process with correlation parameter β; that is, $u(t)$ is stationary Gaussian and Markovian with $E(u(s)u(t)) = \exp\,(-\beta\,|\,t - s\,|)$. Then from the known correspondence (cf. Doob [5])

$$X(t) = \sqrt{t}\, u\left(\frac{1}{2\beta} \log t\right)$$

we obtain

$$b_2(z) = \Pr\left\{ |\, u(t)\,| \leqq z,\ \frac{1}{2\beta} \log \frac{a}{1-a} < t < \frac{1}{2\beta} \log \frac{1}{1-b} \right\},$$

or since the process is strictly stationary

$$b_2(z) = \Pr\left\{ |\, u(t)\,| \leqq z, 0 < t < \frac{1}{2\beta} \log \frac{b(1-a)}{a(1-b)} \right\},$$

which is an absorption probability with a uniform barrier.

The function $b_2(z)$ is of some importance in the theories of communications and statistical equilibrium (cf. Bellman and Harris [1]), and may eventually be tabulated. It seems very difficult to give a complete analysis, but the following partial result is given without proof.

Let $\alpha = \frac{1}{2} \log\,(b(1 - a))/(a(1 - b))$ so that $b_2(z)$ is a function of α. Then it is possible to find the Laplace transform of $b_2(z)$ in the following form:

$$\int_0^\infty e^{-\lambda\alpha} b_2(z)\, d\alpha = \frac{1}{\lambda}\left\{1 - \sqrt{\frac{2}{\pi}}\, \frac{e^{\frac{1}{2}z^2}}{D_{-\lambda}(\sqrt{2}z) + D_{-\lambda}(-\sqrt{2}z)} \int_0^z e^{-\xi^2} \{D_{-\lambda}(\sqrt{2}\xi) + D_{-\lambda}(-\sqrt{2}\xi)\, d\xi \right\},$$

where $D_n(z)$ is the Weber function [23]. It seems very difficult to get even any qualitative information from this formula.

REFERENCES

[1] R. Bellman and T. Harris, "Recurrence times for the Ehrenfest model," *Pacific Jour. Math.*, Vol. 1 (1951), pp. 179–193.

[2] Harald Cramér, "On the composition of elementary errors," *Skandinavisk Aktuarietidskrift*, Vol. 11 (1928), pp. 13–74, 141–180.

[3] G. Doetsch, "Les équations aux dérivées partielles du type parabolique," *Enseignement Math.*, Vol. 35 (1936), pp. 43–87.

[4] M. D. Donsker, "Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, Vol. 23 (1952), pp. 277–281.

[5] J. L. Doob, "The brownian motion and stochastic equations," *Annals of Mathematics*, Vol. 2 (1942), pp. 351–369.

[6] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[7] R. Fortet, "Les fonctions aléatoires du type de Markoff associées à certaines équations linéaires aux dérivées partielles du type parabolique," *Jour. Math. Pures Appl.*, Vol. 22 (1943), pp. 177–243.

[8] E. Goursat, *Cours d'Analyse Mathématique*, Vol. 3, 2nd ed., Gauthier-Villars, Paris, 1915.

[9] A. Hammerstein, "Über Entwicklungen gegebener Funktionen nach Eigenfunktionen von Randwertaufgaben," *Math. Zeit.*, Vol. 27 (1927), pp. 269–311.

[10] M. Kac and A. J. F. Siegert, "An explicit representation of a stationary Gaussian process," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 438–442.

[11] A. N. Kolmogorov, "Sulla determinazione empirica delle leggi di probabilita," *Giorn. Ist. Ital. Attuari*, Vol. 4 (1933), pp. 1–11.

[12] P. Lévy, *Processus Stochastiques et Mouvement Brownien*, Gauthiers-Villars, Paris, 1948.

[13] G. M. Maniya, "Generalization of the criterion of A. N. Kolmogorov," *Doklady Akad. Nauk SSSR (NS)*, Vol. 69 (1949), pp. 495–497.

[14] G. Pólya, "Remarks on computing the probability integral," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp. 63–78.

[15] N. V. Smirnov, "Sur la distribution de ω^2," *C. R. Acad. Sci. Paris*, Vol. 202 (1936), p. 449.

[16] N. V. Smirnov, "On the distribution of the ω^2 criterion," *Rec. Math.* (*Mat. Sbornik*) (*NS*), Vol. 2 (1937), pp. 973–993.

[17] N. V. Smirnov, "On the deviation of the empirical distribution function," *Rec. Math.* (*Mat. Sbornik*) (*NS*), Vol. 6 (1939), pp. 3–26.

[18] N. V. Smirnov, "Table for estimating goodness of fit of empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 279–281.

[19] R. von Mises, *Wahrscheinlichkeitsrechnung*, Deuticke, Vienna, 1931.

[20] R. von Mises, "Differentiable statistical functions," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 309–348.

[21] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 105–118.

[22] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1922.

[23] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge University Press, 1927.

A GENERALIZATION OF THE NEYMAN-PEARSON FUNDAMENTAL LEMMA[1]

By Herman Chernoff and Henry Scheffé

University of Illinois and Columbia University

1. Summary. Given $m + n$ real integrable functions $f_1, \cdots, f_m, g_1, \cdots, g_n$ of a point x in a Euclidean space X, a real function $\phi(z_1, \cdots, z_n)$ of n real variables, and m constants $c_1, \cdots, c_m$, the problem considered is the existence of a set S^0 in X maximizing $\phi\left(\int_S g_1\, dx, \cdots, \int_S g_n\, dx\right)$ subject to the m side conditions $\int_S f_i\, dx = c_i$, and the derivation of necessary conditions and of sufficient conditions on S^0. In some applications the point with coordinates $\left(\int_S g_1\, dx, \cdots, \int_S g_n\, dx\right)$ may also be required to lie in a given set. The results obtained are illustrated with an example of statistical interest. There is some discussion of the computational problem of finding the maximizing S^0.

2. The problem. The Neyman-Pearson fundamental lemma concerns the problem, given a number of integrable functions, to form their integrals over a variable set S, and to find a set S^0 (if any) for which one of these integrals is maximum subject to the condition that the others have fixed values. The generalization considered here is to maximize a function of several integrals, subject to similar side conditions.

More precisely, we are given $m + n$ integrable[2] functions $f_1(x), \cdots, f_m(x)$, $g_1(x), \cdots, g_n(x)$ of a point x in a Euclidean space S, a real-valued function $\phi(z_1, \cdots, z_n)$ of n real variables defined on the n-dimensional Euclidean space Z, or at least on a suitable subset of Z to be specified later, m constants $c_1, \cdots, c_m$, and a subset A of Z. Let S denote any Borel set in X and form

$$\phi\left(\int_S g_1 dx, \cdots, \int_S g_n dx\right). \tag{2.1}$$

The problem is the existence and characterization of sets S^0 which maximize (2.1) subject to the m conditions

$$\int_S f_i dx = c_i \qquad (i = 1, \cdots, m) \tag{2.2}$$

and the further condition that the point with the coordinates

$$\left(\int_S g_1 dx, \cdots, \int_S g_n dx\right)$$

[1] Work of Scheffé sponsored by the Office of Naval Research.

[2] With respect to Lebesgue measure on the Borel sets.

lie in the given subset A of Z. If the only side conditions are of the form (2.2) then $A = Z$. The Neyman-Pearson lemma refers to the case where

$$n = 1, \qquad \phi \equiv z_1, \qquad A = Z. \tag{2.3}$$

Briefly the history of the problem is the following. It arises in the Neyman-Pearson theory of optimum statistical tests and its generalizations. It was treated in the important special case (2.3) by Neyman and Pearson [5], [6], who obtained the inequalities (4.1) below, with the symbol X_c in (4.1) replaced by X, as sufficient conditions for a maximizing set S^0. The problems of existence and necessity in the case (2.3) were recently solved in all generality by Dantzig and Wald [1]; the necessity problem in this case had been solved under some restrictions (including $m = 1$) in the original paper [5] of Neyman and Pearson. A statistical example which does not come under the special case was recently investigated by Isaacson [4], who obtained sufficient conditions for his problem; this example falls under our treatment and is discussed in Section 8.

In this paper we obtain an existence theorem, and necessary conditions and sufficient conditions for a maximizing S^0. To obtain these the results of Dantzig and Wald are employed, as well as their device of considering certain vector measures to which the Lyapunov theorem [3] may be applied. Construction and computation of a maximizing S^0 are also considered.

3. Further notation and the condition $\mathcal{C}$. The symbol S (with or without superscripts) will always be understood to denote a Borel set in the Euclidean space X, and the symbols f_j, g_i will always denote integrable[2] functions of a point x in X. In addition to the n-dimensional Euclidean space Z of points $z = (z_1, \cdots, z_n)$, it is convenient to introduce an m-dimensional Euclidean space Y of points $y = (y_1, \cdots, y_m)$. Furthermore, $y(S)$ will denote the point in Y with the coordinates

$$y_j(S) = \int_S f_j\,dx \qquad (j = 1, \cdots, m),$$

and $z(S)$ the point in Z with the coordinates

$$z_i(S) = \int_S g_i\,dx \qquad (i = 1, \cdots, n).$$

Let c be the point in Y with the coordinates $(c_1, \cdots, c_m)$. Then the quantity (2.1) to be maximized may be written $\phi(z(S))$, and the side conditions as $y(S) = c$ and $z(S) \varepsilon A$.

Call M the range of the $(m + n)$-dimensional vector measure with components

$$y_1(S), \cdots, y_m(S), z_1(S), \cdots, z_n(S), \tag{3.1}$$

that is, M is the set of all points with coordinates (3.1) generated by the totality of Borel sets S in X. Then M is a set in the $(m + n)$-dimensional space $Y \times Z$. By the Lyapunov theorem [3] M is closed, bounded, and convex. We shall call

M_c the set in Z which is the cross section of M by $y = c$, that is, M_c is the set of all points z for which there exists an S with $z(S) = z$, $y(S) = c$. The projection of M on the space Y will be denoted by N, so N is the set of points y for which there exists an S with $y(S) = y$. Thus the side conditions $y(S) = c$ can be satisfied if and only if $c \varepsilon N$.

From the Lyapunov theorem N is a closed, bounded, and convex set in Y. Let π be the smallest-dimensional linear space containing N. In the following it is crucial whether the given point c is an inner point or boundary point of N with respect to the topology not of Y but of π. A point y of N is called an *inner point* of N if there exists an m-dimensional neighborhood U of y such that $U \cap \pi \subset N$, otherwise it is called a *boundary point*.

If c is a boundary point the inequalities of the Neyman-Pearson lemma and its generalizations have to be considered in a certain subset X_c of X determined by the following definitions. Regard the points in Y as vectors and denote the inner product of two vectors $\xi = (\xi_1, \cdots, \xi_m)$ and $y = (y_1, \cdots, y_m)$ by $\xi \cdot y = \xi_1 y_1 + \cdots + \xi_m y_m$. Then $(\xi^1, \cdots, \xi^r)$ is called a *maximal* set of vectors relative to a boundary point c if

$$(3.2) \qquad \xi^i \cdot \xi^i \neq 0 \qquad (i = 1, \cdots, r),$$

$$(3.3) \qquad \xi^1 \cdot y \leqq \xi^1 \cdot c \qquad \text{for all } y \varepsilon N,$$

$$(3.4) \qquad \xi^p \cdot y \leqq \xi^p \cdot c \qquad \begin{array}{l} \text{for all } y \varepsilon N \text{ for which } \xi^i \cdot y = \xi^i \cdot c, \\ i = 1, \cdots, p-1;\ p = 2, \cdots, r. \end{array}$$

A maximal set $(\xi^1, \cdots, \xi^r)$ is called a *complete* maximal set relative to c if $(\xi^1, \cdots, \xi^r, \xi^{r+1})$ maximal relative to c implies ξ^{r+1} is a linear combination of $\xi^1, \cdots, \xi^r$. The existence of a complete maximal set relative to every boundary point c is shown by Dantzig and Wald ([1], Lemma 3.1). The set X_c is now defined as X if c is an inner point, while if c is a boundary point it is defined as the subset of X in which

$$(3.5) \qquad \sum_{j=1}^{m} \xi_j^i f_j(x) = 0 \qquad (i = 1, \cdots, r),$$

where $(\xi^1, \cdots, \xi^r)$ is a complete maximal set relative to c, and $\xi^i = (\xi_1^i, \cdots, \xi_m^i)$.

If D is the domain of definition of $\phi(z)$ and $\phi(z)$ has a differential at $z^0 = (z_1^0, \cdots, z_n^0) \varepsilon D$, then by definition there exist constants[3] $a_1, \cdots, a_n$ such that for $z \varepsilon D$

$$(3.6) \qquad \phi(z) - \phi(z^0) = \sum_{i=1}^{n} a_i(z_i - z_i^0) + o(\| z - z^0 \|),$$

[3] If furthermore z^0 is an interior point of D, then $\partial\phi/\partial z_i$ exist at z^0 and equal a_i $(i = 1, \ldots, n)$. In the converse direction, if $\partial\phi/\partial z_i$ exist in a neighborhood of z^0 and are continuous at z^0, then $\phi(z)$ has a differential at z^0. However, in the applications below z^0 may be a boundary point of D.

where

$$\| z - z^0 \| = \left[\sum_{i=1}^{n} (z_i - z_i^0)^2 \right]^{1/2}.$$

If S^0 is a set such that $\phi(z)$ has a differential at $z^0 = z(S^0)$, and if $a_1, \cdots, a_n$ denote the constants in the differential at z^0 as in (3.6), then S^0 will be said to satisfy the *condition* $\mathcal{C}$ if there exist constants $k_1, \cdots, k_m$ such that

$$\text{(3.7)} \qquad \begin{aligned} \sum_{i=1}^{n} a_i g_i(x) &\geqq \sum_{j=1}^{m} k_j f_j(x) && \text{a.e. in } X_c \cap S^0, \\ \sum_{i=1}^{n} a_i g_i(x) &\leqq \sum_{j=1}^{m} k_j f_j(x) && \text{a.e. in } X_c - S^0. \end{aligned}$$

It should be noted that in general the set S^0 must first be known before the constants a_i can be evaluated. This makes the problem of constructing sets S satisfying the condition $\mathcal{C}$ and the side conditions inherently more difficult in the general case than in the special case (2.3), since in the general case the coefficients a_i in the condition $\mathcal{C}$ are functions of the set S to be determined, while in the special case there is only one a_i which is always unity. Further attention is given to the problem of construction and computation of a maximizing S in Section 9.

About the condition $\mathcal{C}$ we remark also that if c is a boundary point, X_c will frequently be a set of measure zero, in which case the condition $\mathcal{C}$ is vacuous. In this case it may be shown ([1], Lemma 3.2) that the set S satisfying the side conditions $y(S) = c$ is unique up to a set of measure zero.

4. Results of Dantzig and Wald. These concern the special case (2.3). They include an existence theorem which we shall not need (it is covered by our Theorem 5.1) and the following theorem which we shall. Here we write $g_1(x) = g(x)$. The set X_c is defined in connection with (3.5).

THEOREM 4.1 (Dantzig and Wald). *If S^0 is a set satisfying $y(S^0) = c$, then a necessary and sufficient condition that S^0 maximize $\int_S g(s)\,dx$ subject to the condition $y(S) = c$ is that there exist constants $k_1, \cdots, k_m$ such that*

$$\text{(4.1)} \qquad \begin{aligned} g(x) &\geqq \sum_{i=1}^{m} k_i f_i(x) && \text{a.e. } in\ X_c \cap S^0, \\ g(x) &\leqq \sum_{i=1}^{m} k_i f_i(x) && \text{a.e. } in\ X_c - S^0. \end{aligned}$$

5. Existence theorem. For our method of proof of the existence theorem to succeed it is essential that the set A be closed. It may nevertheless be possible to use the theorem in situations where the given A is not closed, by applying it to a closed set A_1 containing A and then arguing that the maximum cannot occur in $A_1 - A$; an example is given in Section 8.

THEOREM 5.1. *If there exists a set S satisfying the side conditions $y(S) = c$ and $z(S) \varepsilon A$, if $\phi(z)$ is continuous in $M_c \cap A$, and if A is closed,*[4] *then there exists a set S^0 maximizing $\phi(z(S))$ subject to the conditions $y(S) = c$ and $z(S) \varepsilon A$.*

PROOF. Since M is closed and bounded, so is M_c, and therefore $M_c \cap A$. Also $M_c \cap A$ is nonempty because there exists an S satisfying $y(S) = c$ and $z(S) \varepsilon A$. Since $\phi(z)$ is continuous in the nonempty closed bounded set $M_c \cap A$, there exists a point $z^0 \varepsilon M_c \cap A$ such that

$$\phi(z^0) = \sup \phi(z) \qquad \text{for } z \varepsilon M_c \cap A.$$

Now $z^0 \varepsilon M_c$ implies the existence of an S^0 with $z(S^0) = z^0$ and $y(S^0) = c$. For any other S satisfying $y(S) = c$ and $z(S) \varepsilon A$ we have $z(S) \varepsilon M_c \cap A$, hence $\phi(z(S)) \leqq \phi(z^0) = \phi(z(S^0))$.

6. Necessary conditions. Suppose $\phi(z)$ takes on its maximum value in $M_c \cap A$ at $z^0 = z(S^0)$. The hypotheses of the following theorem imply that z^0 is an interior (in the topology of the n-dimensional space Z) point of A. This will of course be the case if A is open, and in particular if $A = Z$. On the other hand it is easily seen that z^0 must be a boundary (same topology) point of $M_c \cap A$, unless all the constants a_i (see equation (3.6)) in the differential of $\phi(z)$ at z^0 vanish. An S^0 for which all $a_i = 0$ at $z^0 = z(S^0)$ will always satisfy the condition $\mathfrak{C}$ (with all $k_j = 0$).

THEOREM 6.1. *If S^0 is a set for which $z(S^0)$ is an interior point*[5] *of A, if $\phi(z)$ is defined in $M_c \cap A$ and has a differential at $z = z(S^0)$, then a necessary condition that S^0 maximize $\phi(z(S))$ subject to the conditions $y(S) = c$ and $z(S) \varepsilon A$ is that S^0 satisfy the condition $\mathfrak{C}$.*

PROOF. Assume S^0 satisfies $y(S) = c$ and $z(S) \varepsilon A$, and maximizes $\phi(z(S))$ subject to these conditions. Let $z^0 = z(S^0) = (z_1^0, \cdots, z_n^0)$, let a_i be the constants in the differential of $\phi(z)$ at z^0 as in (3.6), and define

$$\hat{\phi}(z) = a_0 + \sum_{i=1}^{n} a_i z_i,$$

where

$$a_0 = \phi(z^0) - \sum_{i=1}^{n} a_i z_i^0.$$

Then

$$\hat{\phi}(z(S)) = a_0 + \sum_{i=1}^{n} a_i z_i(S) = a_0 + \sum_{i=1}^{n} a_i \int_S g_i \, dx,$$

$$\hat{\phi}(z(S)) = a_0 + \int_S \left(\sum_{i=1}^{n} a_i g_i \right) dx. \tag{6.1}$$

---

[4] A hypothesis of Theorem 5.1 is that there exists a set S^1 satisfying the side conditions. Let $z^1 = z(S^1)$. The hypothesis that A is closed may be replaced by the sometimes useful weaker hypothesis that the set $\{z \mid z \varepsilon (M_c \cap A) \text{ and } \phi(z) \geqq \phi(z^1)\}$ is closed.

[5] The proof shows that this hypothesis may be replaced by the weaker one that $z^0 = z(S^0)$ is a limit point of $L \cap A$ for every line L in Z through z^0.

It will suffice to prove that S^0 maximizes $\hat{\phi}(z(S))$ subject to $y(S) = c$. If this is true we can apply the necessary condition of Theorem 4.1 to $g(x) = \sum_{i=1}^{n} a_i g_i(x)$, and this necessary condition becomes our condition $\mathfrak{C}$.

Suppose then that S^0 does not maximize $\hat{\phi}(z(S))$ subject to $y(S) = c$. Then there exists an S^1 with $y(S^1) = c$ and $\hat{\phi}(z(S^1)) > \hat{\phi}(z(S^0))$. We note that $z^1 = z(S^1)$ is in M_c but not necessarily in A, and that $z^1 \neq z^0$ since $\hat{\phi}(z^1) > \hat{\phi}(z^0)$. Let $\rho = \| z^1 - z^0 \|$, and $h\rho = \hat{\phi}(z^1) - \hat{\phi}(z^0)$, so that $h > 0$. Write

$$z^\lambda = (1 - \lambda)z^0 + \lambda z^1 \qquad (0 \leqq \lambda \leqq 1).$$

Then all $z^\lambda \varepsilon M_c$ since M_c is convex. Because $\hat{\phi}(z)$ is a linear function of z with $\hat{\phi}(z^1) = \hat{\phi}(z^0) + h\rho$, it follows that

$$\hat{\phi}(z^\lambda) = \hat{\phi}(z^0) + \lambda h\rho. \tag{6.2}$$

From (3.6) we have for $z \varepsilon M_c \cap A$

$$\phi(z) = \hat{\phi}(z) + o(\| z - z^0 \|),$$

and hence if $z^\lambda \varepsilon M_c \cap A$,

$$\phi(z^\lambda) = \hat{\phi}(z^\lambda) + o(\lambda\rho).$$

Thus there exists a $\delta > 0$ such that $0 < \lambda\rho < \delta$ and $z^\lambda \varepsilon M_c \cap A$ imply

$$|[\phi(z^\lambda) - \hat{\phi}(z^\lambda)]/(\lambda\rho)| < h,$$

and so

$$\phi(z^\lambda) > \hat{\phi}(z^\lambda) - \lambda h\rho.$$

From this, (6.2), and $\hat{\phi}(z^0) = \phi(z^0)$, we get

$$\phi(z^\lambda) > \phi(z^0) \tag{6.3}$$

if $0 < \lambda < \delta/\rho$ and $z^\lambda \varepsilon M_c \cap A$. Recalling that z^0 is an interior point of A, we see there is a λ', $0 < \lambda' < \delta/\rho$, such that $z^{\lambda'} \varepsilon A$. Also $z^{\lambda'} \varepsilon M_c$, so $z^{\lambda'} \varepsilon M_c \cap A$, and (6.3) is true for $\lambda = \lambda'$. But $z^{\lambda'} \varepsilon M_c \cap A$ also implies that there exists an $S^{\lambda'}$ with $y(S^{\lambda'}) = c$, $z(S^{\lambda'}) = z^{\lambda'} \varepsilon A$. For this $S^{\lambda'}$ we have $\phi(z(S^{\lambda'})) > \phi(z(S^0))$, so S^0 does not maximize $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$. This is a contradiction, and hence S^0 maximizes $\hat{\phi}(z(S))$ subject to $y(S) = c$.

7. Sufficient conditions. It is convenient to introduce a weakened form of the property of concavity of a function, which we shall call quasi-concavity; related concepts have been considered by de Finetti [2]. A function $\phi(z)$ defined in a convex set D is said to be *concave* in D if $z^0 \varepsilon D$, $z^1 \varepsilon D$, $z^\lambda = (1 - \lambda)z^0 + \lambda z^1$, and $0 \leqq \lambda \leqq 1$ imply

$$\phi(z^\lambda) \geqq (1 - \lambda)\phi(z^0) + \lambda\phi(z^1).$$

If D is open and convex, and $\phi(z)$ has continuous second partial derivatives in D, then a necessary and sufficient condition for $\phi(z)$ to be concave in D is that the $n \times n$ matrix

$$(\partial^2\phi/\partial z_i \partial z_j) \tag{7.1}$$

be nonpositive in D, that is, all the characteristic roots be nonpositive in D. We shall say $\phi(z)$ is *quasi-concave* in a convex set D if there exists a real differentiable function $\psi(\phi)$ on an interval I containing the range $\phi(D)$ of $\phi(z)$, with $0 < \psi'(\phi) < +\infty$ for $\phi \varepsilon I$, and such that $\psi(\phi(z))$ is concave in D. We note that concavity implies quasi-concavity (take $\psi(\phi) \equiv \phi$), but not conversely (for example, with $n = 2$ consider $\phi(z) = z_1 z_2$ in the set D where $z_1 > 0$, $z_2 > 0$, and take $\psi(\phi) = \log \phi$).

THEOREM 7.1. *If the set S^0 satisfies the side conditions $y(S) = c$ and $z(S) \varepsilon A$, if $\phi(z)$ is defined and quasi-concave in a convex set containing $M_c \cap A$ and has a differential at $z = z(S^0)$, then a sufficient condition that S^0 maximize $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$ is that S^0 satisfy the condition $\mathfrak{C}$.*

PROOF. Suppose first that $\phi(z)$ is concave instead of merely quasi-concave in a convex set $D \supset M_c \cap A$, that the other hypotheses of the theorem are satisfied by $\phi(z)$ and S^0, and that S^0 satisfies the condition $\mathfrak{C}$. Write $z^0 = z(S^0)$ and define the linear function $\hat{\phi}(z)$ as in the proof of Theorem 6.1. Then S^0 maximizes $\hat{\phi}(z(S))$ subject to the condition $y(S) = c$, since the condition $\mathfrak{C}$ now becomes the sufficient condition of Theorem 4.1 applied to $\hat{\phi}(z(S))$ in the form (6.1).

Next we note that $\phi(z) \leqq \hat{\phi}(z)$ in D. Assume the contrary, that there exists a point $z^1 \varepsilon D$ with $b = \phi(z^1) - \hat{\phi}(z^1) > 0$, so $z^1 \neq z^0$ since $\phi(z^0) = \hat{\phi}(z^0)$. If $z^\lambda = (1 - \lambda)z^0 + \lambda z^1 (0 \leqq \lambda \leqq 1)$, then $z^\lambda \varepsilon D$. Define $\bar{\phi}(z^\lambda) = (1 - \lambda)\phi(z^0) + \lambda\phi(z^1)$. Then $\phi(z^\lambda) \geqq \bar{\phi}(z^\lambda)$ since $\phi(z)$ is concave. But $\hat{\phi}(z^\lambda) = \bar{\phi}(z^\lambda) - \lambda b$, and hence, since $\phi(z)$ has a differential at z^0, $\phi(z^\lambda) < \hat{\phi}(z^\lambda) + \lambda b = \bar{\phi}(z^\lambda)$ for λ sufficiently small but positive. This contradicts $\phi(z^\lambda) \geqq \bar{\phi}(z^\lambda)$.

If now S is any set satisfying $y(S) = c$ and $z(S) \varepsilon A$, then $z(S) \varepsilon M_c \cap A \subset D$, hence $\phi(z(S)) \leqq \hat{\phi}(z(S)) \leqq \hat{\phi}(z(S^0)) = \phi(z(S^0))$, the second inequality because S^0 maximizes $\hat{\phi}(z(S))$ subject to $y(S) = c$. The theorem is now proved in the case where $\phi(z)$ is concave.

Suppose next $\phi(z)$ is quasi-concave in D. By definition there exists a differentiable function $\psi(\phi)$ on an interval I containing $\phi(D)$, such that $0 < \psi'(\phi) < +\infty$ for $\phi \varepsilon I$, and $\Phi(z) \equiv \psi(\phi(z))$ is concave in D. Since $\psi(\phi)$ is a strictly increasing function on I, a set S^0 maximizes $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$ if and only if it maximizes $\Phi(z(S))$ subject to the same side conditions. Since $\Phi(z)$ is concave in D we may apply the above result to $\Phi(z)$ after we verify that $\Phi(z)$ has a differential at $z^0 = z(S^0)$. But this is the case since $\phi(z)$ has a differential at z^0 and $\psi'(\phi)$ exists at $\phi = \phi(z^0)$. Let $\gamma = \psi'(\phi(z^0))$. Then the constants a_i in the differential of $\Phi(z)$ at z^0 are equal to γ times those for $\phi(z)$ at z^0. The factor γ can be absorbed into the constants $k_1, \cdots, k_m$ of the condition $\mathfrak{C}$ since $0 < \gamma < +\infty$.

The following corollary may be useful in applications where it is easier to prove that $\phi(z)$ is suitably dominated by a quasi-concave function than that $\phi(z)$ is quasi-concave.

COROLLARY 7.1. *If the set S^0 satisfies the side conditions $y(S) = c$ and $z(S) \varepsilon A$, if U is a neighborhood in Z of $z^0 = z(S^0)$, if $\phi(z)$ is defined in a set $D \supset U \cup (M_c \cap$*

A) and has a differential at z^0, if $\phi^*(z)$ is defined and quasi-concave in a convex set $D^* \supset D$, and if $\phi(z) \leqq \phi^*(z)$ in D while $\phi(z^0) = \phi^*(z^0)$, then a sufficient condition that S^0 maximize $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$ is that S^0 satisfy the condition $\mathfrak{C}$.

PROOF. The corollary will be an immediate consequence of applying Theorem 7.1 to $\phi^*(z)$ instead of $\phi(z)$, providing we can prove that under the hypotheses of the corollary $\phi^*(z)$ has a differential at z^0, and that this is the same as the differential of $\phi(z)$ at z^0 (else a set of constants $a_1^*, \cdots, a_n^*$ different from $a_1, \cdots, a_n$ would appear in the condition $\mathfrak{C}$).

Suppose $\Phi^*(z) = \psi(\phi^*(z))$ is concave in D^*, where $0 < \psi(\phi) < +\infty$ for $\phi \; \varepsilon \; I \supset \phi^*(D^*)$. Let $\Phi(z) = \psi(\phi(z))$. Since $\psi(\phi)$ has a single-valued differentiable inverse ψ^{-1} on the interval $J = \psi(I)$, and $\phi^*(z) = \psi^{-1}(\Phi^*(z))$, $\phi(z) = \psi^{-1}(\Phi(z))$, it will suffice to prove that $\Phi^*(z)$ has a differential at z^0 and that this is the same as the differential of $\Phi(z)$ at z^0. Since $\Phi^*(z)$ is concave it has a plane of support $w = \bar{\Phi}(z)$ at z^0, that is, there exists a linear function $\bar{\Phi}(z)$ such that $\Phi^*(z) \leqq \bar{\Phi}(z)$ for $z \; \varepsilon \; D^*$ and $\Phi^*(z^0) = \bar{\Phi}(z^0)$. We observe next that this plane is identical with the tangent plane $w = \hat{\Phi}(z)$ to the surface $w = \Phi(z)$ at z^0. For, suppose the contrary. Then because $\bar{\Phi}(z)$ and $\hat{\Phi}(z)$ are both linear and $\bar{\Phi}(z^0) = \hat{\Phi}(z^0)$, there must exist a point z^1 where $\bar{\Phi}(z^1) < \hat{\Phi}(z^1)$. With $z^\lambda = (1 - \lambda)z^0 + \lambda z^1$, $\bar{\Phi}(z^\lambda) < \hat{\Phi}(z^\lambda)$ for $\lambda > 0$. Therefore, since $\Phi(z)$ has a differential at z^0, $\Phi(z^\lambda) > \bar{\Phi}(z^\lambda)$ for λ sufficiently small and positive. But this implies the contradiction $\Phi(z^\lambda) > \Phi^*(z^\lambda)$. From the relation $\Phi(z) \leqq \Phi^*(z) \leqq \bar{\Phi}(z)$ in D, the desired conclusion about the differential of $\Phi^*(z)$ at z^0 easily follows.

8. An example. We will illustrate our results by considering their application in the theory of Type D critical regions for testing simple hypotheses concerning several parameters. Type D regions were recently defined and studied by Isaacson [4]; they are locally optimum unbiased critical regions which are a generalization of the Type A regions of the Neyman-Pearson theory for the one-parameter case.

Suppose X is the sample space and there exists a probability density $p(x, \theta)$ for $\theta = (\theta_1, \cdots, \theta_k)$ in the parameter space Ω. The hypothesis to be tested is is $H_0 : \theta = \theta^0$. We assume that for any set S in X the integral $\int_S p \, dx$ has second partial derivatives with respect to θ_i and $\theta_j (i, j = 1, \cdots, k)$ in a neighborhood of θ^0 which are continuous at θ^0, and that it can be differentiated twice under the integral sign with respect to θ_i and θ_j at θ^0. Denote by $G(S)$ the symmetric matrix $\left(\int_S g_{ij} dx\right)$, where $g_{ij} = [\partial^2 p / \partial\theta_i \partial\theta_j]_{\theta^0}$. Also write

$$f_j = [\partial p / \partial\theta_j]_{\theta^0} \qquad (j = 1, \cdots, k), \; m = k + 1, f_m = p(x, \theta^0). \tag{8.1}$$

It is convenient to call a critical region S for testing H_0 *locally unbiased of size* α if

$$\text{(8.2)} \qquad \begin{aligned} &\int_S f_j\,dx = 0 && (j = 1, \cdots, m-1), \\ &\int_S f_m\,dx = \alpha, \qquad G(S) \text{ is positive definite.} \end{aligned}$$

If S is locally unbiased of size α the (generalized) Gaussian curvature of the power surface at $\theta = \theta^0$ is the determinant $|G(S)|$. A critical region S^0 is said to be *of Type D* if it maximizes $|G(S)|$ subject to the condition that it be locally unbiased of size α. If S^0 is locally unbiased of size α, Isaacson obtained as a sufficient condition for S^0 to be of Type D the existence of constants $k_1, \cdots, k_m$ such that S^0 satisfies

$$\text{(8.3)} \qquad \begin{aligned} &\sum_{i,j=1}^{k} b_{ij}\,g_{ij}(x) \geqq \sum_{j=1}^{m} k_i f_i(x) && \text{a.e. in } S^0, \\ &\sum_{i,j=1}^{k} b_{ij}\,g_{ij}(x) \leqq \sum_{j=1}^{m} k_i f_i(x) && \text{a.e. in } X - S^0, \end{aligned}$$

where the matrix (b_{ij}) of constants is the adjoint matrix of $G(S^0)$.

To make the problem conform better to our previous notation we introduce an n-dimensional space Z of points z, with $n = \frac{1}{2}k(k+1)$, and write the coordinates of z as

$$(z_{11}, z_{12}, \cdots, z_{1k}, z_{22}, \cdots, z_{2k}, z_{33}, \cdots, z_{kk}).$$

Define $\phi(z)$ to be the determinant of the symmetric matrix (z_{ij}), $\phi(z) = |(z_{ij})|$, where $z_{ji} = z_{ij}$. With $z_{ij}(S) = \int_S g_{ij}dx$, we see the problem is to maximize $\phi(z(S))$ subject to the side conditions (8.2). These may be written $y(S) = c$, where $c = (0, \cdots, 0, \alpha)$, and $z(S) \in A$, where A is the part of Z where the matrix (z_{ij}) is positive definite. Since $\phi(z)$ is a polynomial in the coordinates of z it has a differential everywhere. If we write $z^0 = z(S^0)$ and $a_{ij} = [\partial\phi/\partial z_{ij}]_{z^0}$ $(i \leqq j)$, we find $a_{ii} = b_{ii}$, $a_{ij} = 2b_{ij} (i < j)$, and so the condition $\mathfrak{C}$ for the present problem is Isaacson's stated in connection with (8.3), except that the first inequality of (8.3) is asserted a.e. in $X_c \cap S^0$ and the second a.e. in $X_c - S^0$. However, it will be shown later that for all $\alpha \neq 0$ or 1, c is an inner point as defined in Section 3, so that the set X_c is the whole space X, and Isaacson's condition is thus precisely the condition $\mathfrak{C}$ in this case.

To apply our results we need to note that the set A is open and is contained in a closed set A_1 such that $\phi(z) = 0$ in $A_1 - A$. Let $h_i(z)$ with $i = 1, \cdots, 2^k - 1$, denote the determinants of principal minors of the matrix (z_{ij}); these polynomials in the coordinates of z are continuous functions of z. Since A is the set where all $h_i(z) > 0$, A is open. Let A_1 be the set in Z where all $h_i(z) \geqq 0$; then A_1 is closed. In $A_1 - A$ all $h_i(z) \geqq 0$, some $h_j(z) = 0$. Thus (z_{ij}) is positive but not positive definite in $A_1 - A$, and hence its determinant $\phi(z) = 0$ there.

We shall prove first by application of our existence theorem 5.1 that if there

exists any locally unbiased critical region of size α there exists one of Type D. Suppose then that there exists a critical region S^1 satisfying the side conditions (8.2). Then $\phi(z(S^1)) = |G(S^1)| > 0$. Theorem 5.1 tells us there exists a solution S^0 to the modified problem of maximizing $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A_1$. For the solution S^0 we must have $z(S^0) \varepsilon A$, else $z(S^0) \varepsilon A_1 - A$, $\phi(z(S^0)) = 0 < \phi(z(S^1))$. Thus S^0 maximizes $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$.

That any critical region of Type D necessarily satisfies the condition $\mathfrak{C}$ follows immediately from Theorem 6.1.

That the condition $\mathfrak{C}$ is sufficient for a locally unbiased critical region S^0 of size α to be of Type D may be deduced from Theorem 7.1. All the hypotheses of this theorem will be seen to be satisfied if we show that the set A is convex and the function $\log \phi(z)$ is concave in A. Suppose then that ζ^0 and ζ^1 are any two points of A. It will suffice to prove that $\zeta^\lambda = (1 - \lambda)\zeta^0 + \lambda\zeta^1$ is in A and that

$$\log \phi(\zeta^\lambda) \geqq (1 - \lambda) \log \phi(\zeta^0) + \lambda \log \phi(\zeta^1) \tag{8.4}$$

for all $\lambda(0 \leqq \lambda \leqq 1)$. Let ζ^λ_{ij} be the coordinates of ζ^λ. Then the matrices (ζ^r_{ij}) are positive definite for $r = 0, 1$. There thus exists a real nonsingular matrix H such that both matrices $H'(\zeta^0_{ij})H$ and $H'(\zeta^1_{ij})H$ are diagonal, say $H'(\zeta^r_{ij})H = D^r$, where D^r is a diagonal matrix with positive diagonal elements $d^r_1, \cdots, d^r_k (r = 0, 1)$. Now

$$(\zeta^\lambda_{ij}) = (1 - \lambda)(\zeta^0_{ij}) + \lambda(\zeta^1_{ij}),$$

and so $(\zeta^\lambda_{ij}) = K'D^\lambda K$, where $K = H^{-1}$, and D^λ is a diagonal matrix with ith diagonal element equal to $(1 - \lambda)d^0_i + \lambda d^1_i$. Hence D^λ is positive, and so is (ζ^λ_{ij}); thus ζ^λ is in A. Furthermore,

$$\log \phi(\zeta^\lambda) = 2 \log |K| + \log |D^\lambda|,$$

so to prove (8.4) it is enough to verify that

$$\log |D^\lambda| \geqq (1 - \lambda) \log |D^0| + \lambda \log |D^1|,$$

or that

$$\sum_{i=1}^{k} \log [(1 - \lambda) d^0_i + \lambda d^1_i] \geqq (1 - \lambda) \sum_{i=1}^{k} \log d^0_i + \lambda \sum_{i=1}^{k} \log d^1_i .$$

But this follows from the concavity of the function $\log x$.

We shall conclude by proving that c is an inner point of N in this and similar statistical problems with side conditions of the form

$$\int_S p(x, \theta^0)\, dx = \alpha \qquad (0 < \alpha < 1),$$

$$\left[\frac{\partial}{\partial \theta_i} \int_S p(x, \theta)\, dx\right]_{\theta^0} = 0 \qquad (i = 1, \cdots, k),$$

if the integral $\int_S p(x, \theta)\, dx$ can be differentiated once under the integral sign for all (Borel) sets S at $\theta = \theta^0$. With the notation (8.1), and $y_i(S) = \int_S f_i dx$, N is the set of all $y(S)$ in the m-dimensional space Y with $m = k + 1$. We observe that the set N is symmetrical with respect to the point $(0, \cdots, 0, \frac{1}{2})$, that is if $y = (y_1, \cdots, y_{m-1}, y_m)$ is in N so is $(-y_1, \cdots, -y_{m-1}, 1 - y_m)$. For any $y \in N$ there exists an S such that $y(S) = y$. The point $y(X - S)$ is symmetrically placed with respect to $(0, \cdots, 0, \frac{1}{2})$, since for $i = 1, \cdots, m - 1$,

$$y_i(S) + y_i(X - S) = \left[\frac{\partial}{\partial \theta_i}\left\{\int_S p(x, \theta)\, dx + \int_{X-S} p(x, \theta)\, dx\right\}\right]_{\theta^0} = 0,$$

while $y_m(S) + y_m(X - S) = 1$. On taking S to be the empty set we find the point $y^0 = (0, \cdots, 0, 0)$ in N; by symmetry N contains $y^1 = (0, \cdots, 0, 1)$, and by convexity the line segment L joining y^0 and y^1 and containing $c = (0, \cdots, 0, \alpha)$. Since $0 < \alpha < 1$, c is an inner point of the line segment L. From this and the symmetry of N it may be argued geometrically that c is an inner point of N, but we shall give an analytic proof instead.

We shall suppose now that c is a boundary point of the convex body N and from this derive a contradiction. There exists a linear function $h(y) = h_1 y_1 + \cdots + h_m y_m$ not identically zero in N such that $y = c$ maximizes $h(y)$ for y in N. The maximum value of $h(y)$ is thus $h_m \alpha$, and hence $h(y^1) = h_m \leqq h_m \alpha$ and $h(y^0) = 0 \leqq h_m \alpha$; therefore $h_m = 0$. Since zero is the maximum of $h(y)$ for y in N and $h(y)$ does not vanish identically in N, there exists a set S in the space X such that $h(y(S)) < 0$. But

$$h(y(S)) = \sum_{i=1}^{m-1} h_i y_i(S) = -\sum_{i=1}^{m-1} h_i y_i(X - S) = -h(y(X - S)).$$

Thus $h(y(X - S)) > 0$. But $h(y) \leqq 0$ for y in N. This is the desired contradiction.

9. Remarks on computation of a solution. We have mentioned that the problem of construction of a solution is much more difficult here than in the special case covered by the Neyman-Pearson fundamental lemma. We now sketch a general approach which perhaps might be modified and expanded to a method of numerical computation if desired. The basic idea is that the condition $\mathfrak{C}$ reduces the search for a minimizing set among all Borel sets to that for a minimum of a function of $n + m$ real variables or an equivalent problem.

Denote the $(n + m)$-dimensional vector $(\alpha_1, \cdots, \alpha_n, \kappa_1, \cdots, \kappa_m)$ by $v = (v_1, \cdots, v_{n+m})$, and by $S(v)$ the set $\{x \mid \sum_{i=1}^{n} \alpha_i g_i(x) \geqq \sum_{j=1}^{m} \kappa_j f_j(x)\}$. With $y(S)$ and $z(S)$ defined as before, let $Y(v) = y(S(v))$, $Z(v) = z(S(v))$. If $\phi(z)$ has a differential at $z = Z(v)$, denote the differential coefficients there by $\Phi_1(v), \cdots, \Phi_n(v)$. Let $\delta(z, A)$ be a continuous function of z which is nonnegative and vanishes if and only if $z \in A$ (this implies A is closed): an example is the Euclidean dis-

tance from z to A. We now define three functions of v:

$$D(v) = \delta(Z(v), A),$$

$$E(v) = \sum_{j=1}^{m} [Y_j(v) - c_j]^2,$$

where $Y_j(v)$ are the components of $Y(v)$, and

$$F(v) = \sum_{i=1}^{n} [\Phi_i(v) - v_i]^2.$$

The function $F(v)$ is defined only for v such that $\phi(z)$ has a differential at $z = Z(v)$.

We next make the following simplifying assumptions (which would be lightened if the sketch of our method were expanded):

(i). The conditions of our existence theorem, Theorem 5.1, are satisfied.

(ii). $X_c = X$, that is, c is an inner point of N (see Section 3).

(iii). The set $\left\{x \,\middle|\, \sum_{i=1}^{n} \alpha_i g_i(x) - \sum_{j=1}^{m} \kappa_j f_j(x) = 0\right\} \cap (X - X^0)$,

where

$$X^0 = \{x \mid g_1(x) = \cdots g_n(x) = f_1(x) = \cdots = f_m(x) = 0\},$$

has measure zero for all vectors v with the components $\alpha_1, \cdots, \alpha_n$ not all zero.

(iv). For any solution S^0, $z^0 = z(S^0)$ is an interior point of A, and $\phi(z)$ has a nonzero differential at z^0.

(v). $\phi(z)$ is defined and quasi-concave in a convex set containing $A \cap M_c$.

Under assumption (i) a solution of course exists, and under this set of assumptions it is easy to see from Theorems 6.1 and 7.1 that a necessary and sufficient condition for S^0 to be a solution is that, up to a set of measure zero and a subset of X^0, $S^0 = S(v^0)$, where v^0 is a vector v with not all components $v_1, \cdots, v_n$ zero, and

$$D(v^0) = E(v^0) = F(v^0) = 0.$$

The problem has now been reduced to finding a vector v with $v_1, \cdots, v_n$ not all zero satisfying $D(v) = E(v) = F(v) = 0$. This problem can be formulated in various equivalent ways; one is to minimize $D + E + F$.

An inelegant aspect of the above approach is that if v^0 is a solution of the computational problem, then for any positive λ, $S(\lambda v^0) = S(v^0)$ but λv^0 does not satisfy $F(\lambda v^0) = 0$ unless $\lambda = 1$, that is, λv^0 is no longer a solution of the computational problem but $S(\lambda v^0)$ is still the same solution of our actual variational problem. This situation arises from our having required the components $v_i (i = 1, \cdots, n)$ to be equal to $\Phi_i(v)$, when it is sufficient that they be proportional with a positive constant of proportionality. Such a proportionality holds if and only if the function

$$\bar{F}(v) \equiv \left\{\sum_{i=1}^{n} v_i^2 \sum_{j=1}^{n} [\Phi_j(v)]^2\right\}^{\frac{1}{2}} - \sum_{i=1}^{n} v_i \Phi_i(v)$$

vanishes. The inelegancy could thus be removed by replacing $F(v)$ in the above discussion by $\bar{F}(v)$. The solution of the computational problem could then be normalized by adding one of the conditions

$$\sum_{i=1}^{n} v_i^2 = 1 \quad \text{or} \quad \sum_{i=1}^{n+m} v_i^2 = 1.$$

REFERENCES

[1] G. B. Dantzig and A. Wald, "On the fundamental lemma of Neyman and Pearson," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 87–93.

[2] B. de Finetti, "Sulle stratificazioni convesse," *Ann. Mat. Pura Appl. Ser.* 4, Vol. 30 (1949), pp. 173–183.

[3] P. R. Halmos, "The range of a vector measure," *Bull. Am. Math. Soc.*, Vol. 54 (1948), pp. 416–421.

[4] S. L. Isaacson, "On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 217–234.

[5] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London Ser. A*, Vol. 231 (1933), pp. 289–337.

[6] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1–37.

MAXIMUM LIKELIHOOD ESTIMATION IN TRUNCATED SAMPLES[1]

By Max Halperin[2]

USAF School of Aviation Medicine

1. Summary. In this paper we consider the problem of estimation of parameters from a sample in which only the first r (of n) ordered observations are known. If $r = [qn]$, $0 < q < 1$, it is shown under mild regularity conditions, for the case of one parameter, that estimation of θ by maximum likelihood is best in the sense that $\hat{\theta}$, the maximum likelihood estimate of θ, is

(a) consistent,

(b) asymptotically normally distributed,

(c) of minimum variance for large samples.

A general expression for the variance of the asymptotic distribution of $\hat{\theta}$ is obtained and small sample estimation is considered for some special choices of frequency function. Results for two or more parameters and their proofs are indicated and a possible extension of these results to more general truncation is suggested.

2. Introduction. We suppose we are sampling from a univariate population governed by a probability law, $f(x, \theta)$, $-\infty < x < \infty$, where θ is a single parameter. Our sampling process is assumed to be such that for any sample size, n, we have as sample observations only $x_1, x_2, \cdots, x_r$, the r smallest observations in the sample where r is defined for every n by $r = [qn]$. The notation $[a]$ has the usual meaning of the largest integer contained in a. It is assumed that q is known and $0 < q < 1$. Such a sampling process as defined above could easily arise in an experiment of the life-testing variety.

As a case in point, consider the testing of airplane propeller assemblies in a wind tunnel. The assemblies are quite expensive, costing several thousand dollars each. Furthermore, the test, which consists of increasing the wind velocity in the tunnel and observing the velocity at which each assembly is ruptured is of the destructive type. That is, if an assembly fails, it is not repairable, while if it does not fail, its function is not impaired. Thus, on the basis of budget limitations for testing purposes, it may be desirable to limit the number of assemblies that fail. An obvious solution to this problem is to terminate the testing procedure after a fixed percentage of the propellers in the sample fail. The percentage would be fixed in advance so as to keep the total monetary loss within budgetary restrictions. Supposing that the velocity required to rupture a propeller is a random variable following a continuous probability law, we have a simple example of the type of truncated sampling process described above.

[1] Part of a doctoral thesis presented to the Department of Mathematical Statistics, University of North Carolina. Work on this problem was begun while the author was with the Rand Corporation.

[2] Now at National Heart Institute, Bethesda, Maryland.

The sampling process we have described may be generalized to the case of several parameters and further to the case of several points of truncation, each of the latter being defined as a particular sample percentage point. We do not consider these generalizations in detail in this discussion.

To obtain our results we shall need the following assumptions on $f(x, \theta)$. Not all the assumptions are needed in some of our further discussions, but are listed here for brevity and easy reference.

ASSUMPTION A. *For almost all x, the derivatives*

$$\frac{\partial \log f(x,\theta)}{\partial\theta}, \quad \frac{\partial^2 \log f(x,\theta)}{\partial\theta^2}, \quad \frac{\partial^3 \log f(x,\theta)}{\partial\theta^3}, \tag{2.1}$$

exist for every θ belonging to a nondegenerate interval R.

ASSUMPTION B. *For every θ in R we have,*

$$\left|\frac{\partial f(x,\theta)}{\partial\theta}\right| < F_1(x), \qquad \left|\frac{\partial^2 f(x,\theta)}{\partial\theta^2}\right| < F_2(x),$$

$$\left|\frac{\partial^3 f(x,\theta)}{\partial\theta^3}\right| < F_3(x), \qquad \left|\frac{\partial^3 \log f(x,\theta)}{\partial\theta^3}\right| < H(x),$$

where $F_1(x)$, $F_2(x)$, $F_3(x)$ are integrable over $(-\infty, \infty)$, while $\int_{-\infty}^{\infty} H(x)f(x, \theta)\, dx < M$, where M is independent of θ.

ASSUMPTION C. *For every θ in R*

$$K^2 = \int_{-\infty}^{\lambda} \left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)\, dx + \frac{1}{p}\left(\int_{-\infty}^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta}\, dx\right)^2$$

is greater than zero. Here, if θ_0 is the true value of θ, λ is defined by $q = \int_{-\infty}^{\lambda} f(x, \theta_0)\, dx$. That is, λ is the population $100q$ percentage point.

ASSUMPTION D. *$f(x, \theta)$ is continuous in the neighborhood of $x = \lambda$ and has a continuous derivative in x, $f'(x, \theta)$, while*

$$\frac{\partial \log f(x,\theta)}{\partial\theta}, \quad \frac{\partial^2 \log f(x,\theta)}{\partial\theta^2}, \quad \frac{\partial^3 \log f(x,\theta)}{\partial\theta^3},$$

are continuous in the neighborhood of $x = \lambda$.

Finally, we define regular estimation from a joint frequency function, say $h(x_1, \cdots, x_r, \theta)$, in a manner completely analogous to that of Cramér, ([1], p. 479). That is, we suppose we can transform $x_1, \cdots, x_r$ to new variables θ^*, $\lambda_1, \cdots, \lambda_{r-1}$, (where θ^* estimates θ), in a one-to-one manner so that

$$h(x_1, \cdots, x_r; \theta) \prod_{i=1}^{r} dx_i = g(\theta^*; \theta) m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta) \prod_{i=1}^{r-1} d\lambda_i\, d\theta^*, \tag{2.2}$$

where $g(\theta^*; \theta)$ is the density of the estimate θ^*, while $m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta)$ is the conditional density of $\lambda_1, \cdots, \lambda_{r-1}$, given θ^*. Then, if $\partial h/\partial\theta$, $\partial g/\partial\theta$, $\partial m/\partial\theta$

exist for every θ in R and if

$$\left|\frac{\partial h}{\partial \theta}\right| < H_0(x_1, \cdots, x_r), \quad \left|\frac{\partial g}{\partial \theta}\right| < G_0(\theta^*), \quad \left|\frac{\partial m}{\partial \theta}\right| < M_0(\lambda_1, \cdots, \lambda_{r-1}; \theta^*),$$

where H_0, G_0, $\theta^* G_0$, and M_0 are integrable over the whole space of $(x_1, \cdots, x_r)$, θ^*, θ^*, and $\lambda_1, \cdots, \lambda_{r-1}$, respectively, we shall say we are in a regular estimation case of the continuous type and θ^* will be called a regular estimate of θ.

3. Derivation of results. Since our problem is of prominence in the field of life-testing, it is convenient to use a terminology which stems from this connection. It may be remarked that though it is then implied that our random variable is nonnegative the latter point is in no way critical to our proofs.

Thus, let $f(x, \theta)$, $0 \leqq x < \infty$, be a probability density satisfying Assumptions A–D. We suppose that n individuals, each subject to $f(x, \theta)$ as a death law, have been observed from age zero until $r(= [qn])$ of the group have died at times $x_1, x_2, \cdots, x_r$, $0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_r < \infty$. If we denote the sampling density of $x_1, \cdots, x_r$ by $h(x_1, \cdots, x_r)$, we clearly have

$$h(x_1, \cdots, x_r) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(x_i, \theta)[p(x_r, \theta)]^{n-r}, \tag{3.1}$$

where $p(x_r, \theta) = 1 - q(x_r, \theta) = \int_{x_r}^{\infty} f(x, \theta)\, dx$. If we further denote the conditional joint density of $x_1, \cdots, x_{r-1}$, given x_r, by $h(x_1, \cdots, x_{r-1}; x_r)$ and denote the density of x_r in a sample of n by $S_n(x_r)$, we have

$$\begin{aligned} h(x_1, \cdots, x_r) &= h(x_1, \cdots, x_{r-1}; x_r) S_n(x_r) \\ &= (r-1)! \prod_{i=1}^{r-1} \left[\frac{f(x_i, \theta)}{q(x_r, \theta)}\right] r\binom{n}{r} f(x_r, \theta)[p(x_r, \theta)]^{n-r}[q(x_r, \theta)]^{r-1}. \end{aligned} \tag{3.2}$$

We have now, denoting by the symbol E the operation of taking an expected value, the following lemma

LEMMA 1. *If Assumptions A and B hold,*

$$E\left[\frac{\partial \log h(x_1, \cdots, x_r)}{\partial \theta}\right]^2 = -E\left[\frac{\partial^2 \log h(x_1, \cdots, x_r)}{\partial \theta^2}\right]. \tag{3.3}$$

PROOF. The proof consists of verifying that under Assumptions A–D

$$\int_{E_r} \frac{\partial h(x_1, \cdots, x_r)}{\partial \theta} \prod_{i=1}^{r} dx_i = \int_{E_r} \frac{\partial^2 h(x_1, \cdots, x_r)}{\partial \theta^2} \prod_{i=1}^{r} dx_i = 0, \tag{3.3.1}$$

and then proceeding exactly as in Cramér ([1], p. 502). Here E_r is the domain of $x_1, \cdots, x_r$.

In order for (3.3.1) to hold we must have $|\partial h/\partial \theta| < H_0(x_1, \cdots, x_r)$, $|\partial^2 h/\partial \theta^2| < H_1(x_1, \cdots, x_r)$, where H_0 and H_1 are integrable over E_r. We have

$$\frac{\partial h}{\partial \theta} = \frac{n!}{(n-r)!} \sum_{i=1}^{r} \prod_{j \neq i} f(x_j, \theta) \frac{\partial f(x_i, \theta)}{\partial \theta} [p(x_r, \theta)]^{n-r} + \frac{n!}{(n-r-1)!} \prod_{i=1}^{r} f(x_i, \theta)[p(x_r, \theta)]^{n-r-1} \left[\frac{\partial p(x_r, \theta)}{\partial \theta}\right],$$

and

$$\left|\frac{\partial h}{\partial \theta}\right| < \frac{n!}{(n-r)!} \sum_{i=1}^{r} \prod_{j \neq i} f(x_j, \theta) F_1(x_i) + \frac{n!}{(n-r-1)!} \prod_{i=1}^{r} f(x_i, \theta) \int_0^{\infty} F_1(x)\, dx$$

from Assumption B. We also know that $\partial f(x, \theta)/\partial \theta$ exists for all θ in some interval. Thus we may choose a θ_0 in that interval and assert for all θ in the interval

$$f(x, \theta) < f(x, \theta_0) + F_1(x)\, d = F_0(x),$$

where d is the length of the continuity interval on θ. We then have

$$\left|\frac{\partial h}{\partial \theta}\right| < \frac{n!}{(n-r)!} \sum_{i=1}^{r} \prod_{j \neq i} F_0(x_j) F_1(x_i) + \frac{n!}{(n-r-1)!} \prod_{i=1}^{r} F_0(x_i) \int_0^{\infty} F_1(x)\, dx = H_0(x_1, \cdots, x_r).$$

It is clear that $H_0(x_1, \cdots, x_r)$ as just defined is integrable over E_r. A similar discussion holds for $\partial^2 h/\partial \theta^2$ and the lemma follows.

Next we prove the following lemma.

LEMMA 2. *Let $\theta^* = \theta^*(x_1, \cdots, x_r)$ be an unbiased estimate of θ, θ^* being continuous and possessing partial derivatives $\partial \theta^*/\partial x_j (j = 1, 2, \cdots, r)$ in almost all points $(x_1, \cdots, x_r)$. If estimation from $h(x_1, \cdots, x_r)$ is regular, we have asymptotically*

$$nE(\theta^* - \theta)^2 \geq \frac{1}{K^2}, \tag{3.4}$$

where

$$K^2 = \int_0^{\lambda} \left[\frac{\partial \log f(x, \theta)}{\partial \theta}\right]^2 f(x, \theta)\, dx + \frac{1}{p}\left[\int_0^{\lambda} \frac{\partial f(x, \theta)}{\partial \theta}\, dx\right]^2.$$

PROOF. Consider

$$\int_{-\infty}^{\infty} g(\theta^*; \theta)\, d\theta^* = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta) \prod_{i=1}^{r-1} d\lambda_i = 1,$$

where $g(\theta^*; \theta)$ and $m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta)$ are as defined in Section 2, in the definition of a regular estimation case. Under our regularity assumptions on $m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta)$ and $g(\theta^*; \theta)$, these integrals may be differentiated with respect to θ under the integral signs. Thus we have

$$\begin{aligned} &\int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial \theta}\right) g(\theta^*; \theta)\, d\theta^* \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\partial \log m}{\partial \theta}\right) m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta) \prod_{i=1}^{r-1} d\lambda_i = 0. \end{aligned} \tag{3.4.1}$$

Now, referring back to (2.2), we take logarithms of both sides of that relationship, neglecting differentials, differentiate and have

$$\frac{\partial \log h(x_1, \cdots, x_r)}{\partial\theta} = \frac{\partial \log g(\theta^*; \theta)}{\partial\theta} + \frac{\partial \log m}{\partial\theta}. \tag{3.4.2}$$

Squaring both sides of (3.4.2) and taking expected values of each side, we get

$$\begin{aligned} \int_{E_r} \left(\frac{\partial \log h}{\partial\theta}\right)^2 h(x_1, \cdots, x_r) \prod_{i=1}^{r} dx_i &= \int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial\theta}\right)^2 g(\theta^*; \theta)\, d\theta^* \\ &\quad + \int_{-\infty}^{\infty} g(\theta^*; \theta) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\partial m}{\partial\theta}\right)^2 m \prod_{i=1}^{r-1} d\lambda_i\, d\theta^* \\ &\geqq \int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial\theta}\right)^2 g(\theta^*; \theta)\, d\theta^*. \end{aligned} \tag{3.4.3}$$

The cross-product terms, resulting from squaring, all vanish by (3.4.1). Since, under our assumptions, we can write (cf. [1], p. 475), for θ^* unbiased,

$$E(\theta^* - \theta)^2 \geqq \left[\int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial\theta}\right)^2 g(\theta^*; \theta)\, d\theta^*\right]^{-1}, \tag{3.4.4}$$

it follows that we have exactly

$$nE(\theta^* - \theta)^2 \geqq \left\{\frac{1}{n} E\left[\frac{\partial \log h(x_1, \cdots, x_r)}{\partial\theta}\right]^2\right\}^{-1}, \tag{3.4.5}$$

or by (3.3)

$$\geqq -\left\{\frac{1}{n} E \frac{\partial^2 \log h(x_1, \cdots, x_r)}{\partial\theta^2}\right\}^{-1}.$$

From (3.1) we can calculate $(\partial^2 \log h(x_1, \cdots, x_r))/(\partial\theta^2)$ in detail, and integrating out $x_1, \cdots, x_{r-1}$, we get, after some manipulation,

$$\begin{aligned} -\frac{1}{n} E \frac{\partial^2 \log h(x_1, \cdots, x_r)}{\partial\theta^2} &= -\int_0^{\infty} \frac{\partial^2 q(x_{r-1}, \theta)}{\partial\theta^2} S_{n-1}(x_{r-1})\, dx_{r-1} \\ &\quad + \int_0^{\infty} \frac{\partial^2 q(x_r, \theta)}{\partial\theta^2} S_{n-1}(x_r)\, dx_r \\ &\quad + \frac{n-1}{(n-r-1)} \int_0^{\infty} \left[\frac{\partial q(x_r, \theta)}{\partial\theta}\right]^2 S_{n-2}(x_r)\, dx_r \\ &\quad + \int_0^{\infty} \left[\int_0^{x_{r-1}} \left(\frac{\partial \log f(x, \theta)}{\partial\theta}\right)^2 f(x, \theta)\, dx\right] S_{n-1}(x_{r-1})\, dx_{r-1} \\ &\quad - \frac{1}{n} \int_0^{\infty} \frac{\partial^2 \log f(x_r, \theta)}{\partial\theta^2} S_n(x_r)\, dx_r. \end{aligned} \tag{3.4.6}$$

In (3.4.6)

$$S_{n-c}(x_{r-d}) = \frac{(n-c)!}{(r-d-1)!\,(n-c-r+d)!} [q(x_{r-d}, \theta)]^{r-d-1} [p(x_{r-d}, \theta)]^{n-c-r+d} f(x_{r-d}, \theta).$$

That is, $S_{n-c}(x_{r-d})$ is simply the sampling likelihood of the $(r-d)$th smallest order statistic in a sample of size $(n-c)$. With this understanding the basis for the right-hand side of (3.4.6) is readily apparent. If we consider the last term on the right-hand side of (3.4.6), we see that

$$\frac{1}{n}\left|\int_0^\infty \frac{\partial^2 \log f(x_r, \theta)}{\partial\theta^2} S_n(x_r)\, dx_r\right| \leqq \frac{1}{\sqrt{n}}\int_0^\infty \left|\frac{\partial^2 \log f(x, \theta)}{\partial\theta^2}\right| f(x, \theta)\, dx,$$

which is $O(1/\sqrt{n})$, since from our assumptions the integral exists. Hence, asymptotically we can disregard such a term. Now we consider the integrands of the remaining terms of (3.4.6) and on the integrand containing $S_{n-c}(x_{r-d})$, we perform the transformation

$$y = \frac{\sqrt{n-c}}{a}(x_{r-d} - \lambda), \tag{3.4.7}$$

where $a = (\sqrt{qp}/f(\lambda, \theta))$. It follows from Cramér [1], pp. 367–9, that the functions $(a/\sqrt{n-c})S_{n-c}(\lambda + ay/\sqrt{n-c})$ each converge to $1/\sqrt{2\pi}\exp(-\frac{1}{2}y^2)$ in any finite y interval and are each uniformly bounded in any such interval. Denoting the function associated with $S_{n-c}(\lambda + (a/\sqrt{n-c})y)$ by $g_{n-c}(\lambda + (a/\sqrt{n-c})y)$, it is apparent that we can expand g_{n-c} in series about $y = 0$ to zero-order terms plus a remainder, and consequently that

$$\lim_{n\to\infty} g_{n-c}\left(\lambda + \frac{a}{\sqrt{(n-c)}}y\right) = g(\lambda), \text{ say,} \tag{3.4.8}$$

for any fixed y. Furthermore, it is clear that each g_{n-c} is uniformly bounded in every y interval, finite or infinite. Thus the general relation desired is that

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} S_n(y)g_n(y)\, dy = \frac{g}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-\tfrac{1}{2}y^2)\, dy, \tag{3.4.9}$$

where we know that $g_n(y)$ and $S_n(y)$ converge to $g(\lambda)$ and $1/\sqrt{2\pi}\exp(-\frac{1}{2}y^2)$ respectively, for any fixed y, and that $g_n(y)$ is absolutely bounded by a constant, G, while $S_n(y)$ is uniformly bounded in any finite y interval. To establish (3.4.9) we choose a $y_0 > 0$ such that, for any preassigned $\epsilon > 0$

$$\frac{1}{\sqrt{2\pi}}\int_{|y|\leq y_0}\exp(-\tfrac{1}{2}y^2)\, dy = 1 - \frac{\epsilon}{6(G + |g|)}.$$

We can also write

$$\begin{aligned}&\left|\int_{-\infty}^{\infty} g_n(y)S_n(y)dy - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g\exp(-\tfrac{1}{2}y^2)\, dy,\right| \\ &\qquad\leqq \left|\int_{|y|\leq y_0} g_n(y)S_n(y)\, dy - \frac{g}{\sqrt{2\pi}}\int_{|y|\leq y_0} g\exp(-\tfrac{1}{2}y^2)\, dy\right| \\ &\qquad\qquad + G\int_{|y|>y_0} S_n(y)\, dy + \frac{|g|}{\sqrt{2\pi}}\int_{|y|>y_0}\exp(-\tfrac{1}{2}y^2)\, dy.\end{aligned}$$

Since for $|y| \leqq y_0$, $g_n(y)S_n(y)$ converges everywhere to $g/\sqrt{2\pi} \exp(-\frac{1}{2}y^2)$ and since $g_n(y)S_n(y)$ is uniformly bounded in this interval, it follows that we can choose an n_0' such that for $n > n_0'$,

$$\left| \int_{|y| \leq y_0} g_n(y)S_n(y)\, dy - \frac{g}{\sqrt{2\pi}} \int_{|y| \leq y_0} \exp(-\tfrac{1}{2}y^2)\, dy \right| < \tfrac{1}{2}\epsilon.$$

We also have by construction

$$\frac{|g|}{\sqrt{2\pi}} \int_{|y| > y_0} \exp(-\tfrac{1}{2}y^2)\, dy = \frac{|g|\,\epsilon}{6(G + |g|)} < \frac{\epsilon}{6}.$$

Finally, we have

$$G \int_{|y| \geq y_0} S_n(y)\, dy = G\left[1 - \int_{|y| \leq y_0} S_n(y)\, dy\right] > 0,$$

and for $|y| \leqq y_0$, we can choose an n_0'' such that for $n > n_0''$

$$\int_{|y| \leq y_0} S_n(y)\, dy > \frac{1}{\sqrt{2\pi}} \int_{|y| \leq y_0} \exp(-\tfrac{1}{2}y^2)\, dy - \frac{\epsilon}{6G},$$

so that

$$G \int_{|y| \leq y_0} S_n(y)\, dy < G\left[1 - 1 + \frac{\epsilon}{6(G + |g|)} + \frac{\epsilon}{6G}\right] < \frac{\epsilon}{3}.$$

Thus, choosing $n_0 = \max(n_0', n_0'')$, we can assert that for any preassigned $\epsilon > 0$, we can find an n_0 such that for $n > n_0$,

$$\left| \int_{-\infty}^{\infty} g_n(y)S_n(y)\, dy - \frac{g}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}y^2)\, dy \right| < \epsilon.$$

Thus (3.4.9) holds. Then taking the limit of (3.4.6) and simplifying, (3.4) follows.

We are now ready to prove the following theorem.

THEOREM 1. *The likelihood equation*

$$\text{(3.5)} \qquad \frac{\partial \log h(x_1, \cdots, x_r)}{\partial \theta} = \sum_{i=1}^{r} \frac{\partial \log f(x_i, \theta)}{\partial \theta} + (n - r) \frac{\partial \log p(x_r, \theta)}{\partial \theta} = 0,$$

corresponding to (3.1) *has a root* $\hat{\theta}$ *which*

(1) *converges in probability* (*i.p.*) *to the true value of* θ;

(2) *is asymptotically normally distributed*;

(3) *is asymptotically efficient.*

PROOF. First we show $\hat{\theta}$ converges i.p. to the true value of θ, θ_0, say. We can write

$$\text{(3.5.1)} \qquad \begin{aligned} \frac{1}{n} \frac{\partial \log h(x_1, \cdots, x_r)}{\partial \theta} &= \frac{1}{n}\left(\frac{\partial \log h}{\partial \theta}\right)_{\theta_0} + \frac{(\theta - \theta_0)}{n}\left(\frac{\partial^2 \log h}{\partial \theta^2}\right)_{\theta_0} \\ &\quad + \tfrac{1}{2}\Delta(\theta - \theta_0)^2 T(x_1, \cdots, x_r) = B_0 + (\theta - \theta_0)B_1 + \tfrac{1}{2}\Delta(\theta - \theta_0)^2 B_2. \end{aligned}$$

Here $|\Delta| < 1$, the subscript θ_0 denotes evaluation at θ_0, and B_0, B_1, B_2 are

functions of the random variables $x_1, \cdots, x_r$. We note that for the method of proof used here we must have

$$\frac{1}{n}\left|\frac{\partial^3 \log h}{\partial\theta^3}\right| < T(x_1, \cdots, x_r), \tag{3.5.2}$$

where

$$ET(x_1, \cdots, x_r) < M, \tag{3.5.3}$$

where M is a positive constant independent of θ. If we assume, in addition to Assumptions A–D, that $1/p(x_r, \theta)$ is bounded independent of θ, say by $I(x_r)$, where $E[I(x_r)] < I$, where I is independent of θ, (3.5.2) and (3.5.3) are easily seen to hold. The calculations are simple and are omitted.

Now we consider the characteristic function, $\phi_0(t)$ of B_0. We have

$$\begin{aligned}\phi_0(t) = \int_{E_r} \exp\left[\frac{it}{n}\left\{\sum_{i=1}^{r} \frac{\partial \log f(x_i, \theta)}{\partial\theta} - \frac{(n-r)}{p(x_r, \theta)}\int_0^{x_r} \frac{\partial f(x, \theta)}{\partial\theta}\, dx\right\}\right] \\ \cdot\, h(x_1, \cdots, x_r) \prod_{i=1}^{r} dx_i,\end{aligned} \tag{3.5.4}$$

or integrating on $x_1, \cdots, x_{r-1}$,

$$= \int_0^\infty \left[U\left(\frac{t}{n}, x_r\right)\right]^{r-1} V\left(\frac{t}{n}, x_r\right)\left[W\left(\frac{t}{n}, x_r\right)\right]^{n-r} S_n(x_r)\, dx_r,$$

where

$$U\left(\frac{t}{n}, x_r\right) = \frac{\displaystyle\int_0^{x_r} \exp\left\{\frac{it}{n}\frac{\partial \log f(x, \theta)}{\partial\theta}\right\} f(x, \theta)\, dx}{q(x_r, \theta)},$$

$$V\left(\frac{t}{n}, x_r\right) = \exp\left\{\frac{it}{n}\frac{\partial \log f(x_r, \theta)}{\partial\theta}\right\},$$

$$W\left(\frac{t}{n}, x_r\right) = \exp\left\{-\frac{it}{n}\frac{\partial q(x_r, \theta)}{\partial\theta}\frac{1}{p(x_r, \theta)}\right\}.$$

Now on the integrand of (3.5.4), we perform the transformation (3.4.7) and have

$$\begin{aligned}\frac{a}{\sqrt{n}}\left[U\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right)\right]^{r-1} V\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right) \\ \cdot\left[W\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right)\right]^{n-r} S_n\left(\lambda + \frac{ay}{\sqrt{n}}\right).\end{aligned} \tag{3.5.5}$$

We want the limit for fixed y and t of (3.5.5). From [1], pp. 367–69, we have

$$\lim_{n\to\infty} \frac{a}{\sqrt{n}} S_n\left(\lambda + \frac{ay}{\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}y^2\right)$$

for every fixed y. Also

$$\lim_{n\to\infty} \log V\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right) = \lim_{n\to\infty} \frac{it}{n} \frac{\partial \log f\left(\lambda + \frac{ay}{\sqrt{n}}\right)}{\partial\theta} = 0,$$

for every fixed y and t, from the continuity of $\partial \log f(x, \theta)/\partial\theta$ about $x = \lambda$. If we now consider the function $U(t, \lambda + ay)$, we see that for every fixed y it is a characteristic function. Further $(\partial U/\partial t)_{t=0}$ exists for every fixed y. Thus we can expand $U(t, \lambda + ay)$ in a series in t and have for values of t near zero

$$\begin{aligned} U(t, \lambda + ay) &= 1 + \int_0^{\lambda+ay} \frac{\partial f(x,\theta)}{\partial\theta}\,dx \frac{it}{q(\lambda + ay, \theta)} \\ &+ \frac{\left[\int_0^{\lambda+ay} \frac{\partial f(x,\theta)}{\partial\theta} \exp\left\{i\Delta_1 t \frac{\partial \log f(x,\theta)}{\partial\theta}\right\} dx - \int_0^{\lambda+ay} \frac{\partial f(x,\theta)}{\partial\theta}\,dx\right] it}{q(\lambda + ay, \theta)}, \end{aligned} \tag{3.5.6}$$

where $|\Delta_1| < 1$ and the term in square brackets goes to zero with t. We can also write, for example

$$\frac{\int_0^{\lambda+ay} \frac{\partial f(x,\theta)}{\partial\theta}\,dx}{q(\lambda + ay)} = \frac{\int_0^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta}\,dx}{q} + \left[\frac{ay}{q(\lambda + a\Delta_2 y, \theta)} \frac{\partial f(\lambda + a\Delta_2 y, \theta)}{\partial\theta} - \frac{af(\lambda + a\Delta_2 y, \theta) \int_0^{\lambda+a\Delta_2 y} \frac{\partial f(x,\theta)}{\partial\theta}\,dx}{q(\lambda + a\Delta_2 y, \theta)}\right],$$

where $|\Delta_2| < 1$. Then putting t/n and $y/\sqrt{n}$ for t and y respectively, we get

$$U\left(\frac{t}{n} + \frac{ay}{\sqrt{n}}\right) = 1 + \frac{\int_0^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta}\,dx}{q} \frac{it}{n} + \frac{\rho_1(n, t, y)}{n}, \tag{3.5.7}$$

where $\rho_1(n, t, y)$ approaches zero for any fixed y and t, as $n \to \infty$. Similar considerations lead us to

$$W\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right) = 1 - \frac{\int_0^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta}\,dx}{p} \frac{it}{n} + \frac{\rho_2(n, t, y)}{n}, \tag{3.5.8}$$

where $\rho_2(n, t, y)$ approaches zero for any fixed y and t as $n \to \infty$. It follows that (3.5.5) becomes, asymptotically,

$$\exp\left[it\left\{\int_0^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta}\,dx - \int_0^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta}\,dx\right\}\right] \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}y^2\right). \tag{3.5.9}$$

Since (3.5.5) meets the conditions indicated for validity of (3.4.9), we can apply

a convergence argument as indicated in Lemma 2 and conclude

$$\lim_{n\to\infty} \phi_0(t) = \exp\left[it\left\{\int_0^\lambda \frac{\partial f(x,\theta)}{\partial\theta}\,dx - \int_0^\lambda \frac{\partial f(x,\theta)}{\partial\theta}\,dx\right\}\right], \tag{3.5.10}$$

so that B_0 converges i.p. to zero.

Similar arguments lead us to

$$\lim_{n\to\infty} E \exp(itB_1) = \exp(-K^2 it), \tag{3.5.11}$$

and

$$\lim_{n\to\infty} E \exp(itB_2) = \exp(M'it), \tag{3.5.12}$$

so that B_1 converges i.p. to $-K^2$ while B_2 converges i.p. to $M' < M$, a positive constant independent of θ. The precise argument given in [1], pp. 502–3, may then be employed to show that (3.5) has a solution, $\hat\theta$, which converges in probability to θ_0. We omit these arguments.

Now from (3.5.1), we have

$$K\sqrt{n}(\theta - \theta_0) = \frac{\dfrac{1}{K\sqrt{n}}\left(\dfrac{\partial \log h}{\partial\theta}\right)_{\theta_0}}{-\dfrac{B_1}{K^2} - \dfrac{\Delta}{2K^2}B_2(\hat\theta - \theta_0)}. \tag{3.5.13}$$

The denominator of the right-hand side of (3.5.13) converges i.p. to 1, so that we may infer by well known theorems that the asymptotic distribution of the ratio is simply the asymptotic distribution of the numerator. Thus we need only show that $(1/K\sqrt{n})(\partial \log h/\partial\theta)_{\theta_0}$ is asymptotically normal with zero mean and unit variance in order to complete the proof of our theorem. Denoting the characteristic function of $(1/K\sqrt{n})(\partial \log h/\partial\theta)_{\theta_0}$ by $\phi(t)$, we have, by virtue of (3.5.4)

$$\phi(t) = \phi_0\left(\frac{\sqrt{n}\,t}{K}\right). \tag{3.5.14}$$

Applying the transformation (3.4.7) to (3.5.14) we can exactly as before show that $V(t/K\sqrt{n}, \lambda + ay/\sqrt{n})$ converges to 1 for every fixed y and t, while $(a/\sqrt{n})S_n(\lambda + ay/\sqrt{n})$ converges to $(1/\sqrt{2\pi}) \exp -\frac{1}{2}y^2$ for every fixed y. If now we turn our attention to $U(t, \lambda + ay)$, we see that U can be expanded near $t = 0$ in powers of t to terms of order t^2 plus a remainder of order $o(t^2)$, since the second moment of the distribution corresponding to U exists for every fixed y. Similar remarks apply to $W(t, \lambda + ay)$. Thus, by manipulation of the type employed in obtaining an asymptotic representation of (3.5.5), we find that the similar result for the integrand of (3.5.14) is

$$\begin{aligned}\frac{1}{\sqrt{2\pi}} \exp\Bigg[-\tfrac{1}{2}\Bigg\{y^2 &- \frac{2i}{K\sqrt{pq}}\left(\int_0^\lambda \frac{\partial f(x,\theta)}{\partial\theta}\,dx\right) ty \\ &+ \int_0^\lambda \left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)\,dx - \frac{1}{q}\left(\int_0^\lambda \frac{\partial f(x,\theta)}{\partial\theta}\,dx\right)^2 \frac{t^2}{K^2}\Bigg\}\Bigg].\end{aligned} \tag{3.5.15}$$

Since (3.5.14) meets the conditions indicated for validity of (3.4.9), we can for any fixed t, carry out a convergence argument as indicated in Lemma 2, to obtain

$$\lim_{n\to\infty} \phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}t^2\right) \cdot \exp\left[-\frac{1}{2}\left\{y + \frac{it}{K\sqrt{pq}} \int_0^\lambda \frac{\partial f(x,\theta)}{\partial\theta}\,dx\right\}^2\right] dy = \exp\left(-\tfrac{1}{2}t^2\right). \tag{3.5.16}$$

Theorem 1 follows.

4. Generalizations. One can generalize the discussion of Section 3 to the case of several parameters and show that maximum likelihood estimation of $\theta = (\theta_1, \cdots, \theta_p)$ from $h(x_1, \cdots, x_r)$ is a best estimation procedure in the sense that

(a) the maximum likelihood equations have a set of solutions $(\hat\theta_1, \cdots, \hat\theta_p)$ which are consistent;

(b) $\sqrt{n}(\hat\theta - \theta)$ has a multivariate normal limit law;

(c) the covariance matrix of the limiting distribution of $\sqrt{n}(\hat\theta - \theta)$ is the best matrix in the sense of Cramér [1].

In connection with (c) we mean specifically that the concentration ellipsoid corresponding to the covariance matrix of the limiting distribution of $\sqrt{n}(\hat\theta - \theta)$ is identical with

$$\sum_{i,j=1}^{p} E \frac{\partial \log h}{\partial\theta_i} \frac{\partial \log h}{\partial\theta_j} (\mu_i - \theta_i)(\mu_j - \theta_j) = p + 2. \tag{4.2}$$

The ellipsoid (4.2) is shown by Cramér [1] to lie wholly within the ellipsoid corresponding to any set of regular unbiased estimates of $\theta_1, \cdots, \theta_p$. The meaning of "regular" here is precisely in the sense of Cramér [1] as applied to the joint frequency function $h(x_1, \cdots, x_r)$. The assumptions necessary to obtain the result are the natural analogues of Assumptions A–D. Thus A, B, D are extended by imposing similar conditions upon the various derivatives up to third order, that is those with respect to each θ_i and also the mixed derivatives. The condition C becomes a requirement that the matrix with elements

$$A_{ij} = \int_0^\lambda \left(\frac{\partial \log f(x,\theta)}{\partial\theta_i}\right)\left(\frac{\partial \log f(x,\theta)}{\partial\theta_j}\right) f(x,\theta)\,dx + \frac{1}{p}\left(\int_0^\lambda \frac{\partial f(x,\theta)}{\partial\theta_i}\,dx\right)\left(\int_0^\lambda \frac{\partial f(x,\theta)}{\partial\theta_j}\,dx\right), \qquad i, j = 1, 2, \cdots, p,$$

be positive definite. The additional assumption on $p(x_r, \theta)$ specified in Theorem 1 remains unchanged except that θ is taken to be a vector parameter.

Under the assumptions outlined above the proof of (a), (b), (c) follows the lines of Section 3.

A direction of further generalization is to the case of several points of trunca-

tion, each truncation point being a sample percentage point. This work has not been carried out in detail, but due to the asymptotic joint normality of sample percentage points, it appears clear that results of the nature of (a), (b), (c) would hold under conditions analogous to these given by Assumptions A–D.

5. Small-sample estimation. For samples of the type considered in Section 3, one can obtain small-sample results for two important special choices of $f(x, \theta)$.

CASE A. $$f(x, \theta) = \theta e^{-\theta x}, \qquad 0 \leqq x < \infty, \theta > 0.$$

For this case we have from (3.1)

$$h(x_1, \cdots, x_r) = \frac{n!}{(n-r)!} \theta^r \exp\left\{ -\theta \left[\sum_{i=1}^{r-1} x_i + (n - r + 1)x_r \right] \right\}. \tag{5.1}$$

From (3.4) it appears that in any regular estimation case an estimate θ^* will be such that

$$E(\theta^* - \theta)^2 \geqq \frac{\theta^2}{nq}, \tag{5.1.1}$$

for θ^* unbiased. From (5.1) we obtain the maximum likelihood estimate of θ as

$$\hat{\theta} = \frac{r}{\sum_{i=1}^{r-1} x_i + (n - r + 1)x_r} = \frac{r}{y}, \quad \text{say.} \tag{5.1.2}$$

It is easy to show by calculating its moment generating function that the random variable, $2\theta y$, has a chi-square distribution with $2r$ degrees of freedom. It follows that

$$E\hat{\theta} = \frac{r\theta}{r-1} \sim \theta, \qquad \text{Var } \hat{\theta} = \frac{\theta^2}{r-1} \sim \frac{\theta^2}{nq}. \tag{5.1.3}$$

Thus $\hat{\theta}$ is a best estimate in the sense of Theorem 1. $\hat{\theta}$ can, of course, be corrected for bias, the variance of the adjusted estimate then being $\theta^2/(r - 2)$.

CASE B. $$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right], \qquad -\infty < x < \infty.$$

This case is of marked interest since it is frequently assumed in life testing that the logarithm of time to death is normally distributed. Essentially this case has also been considered by Hald [3] and Cohen [5]. We indicate the solution for completeness. It can be shown that

$$\hat{\sigma} = \tfrac{1}{2}(x_r - \bar{x}_r)(-\hat{h} + \sqrt{\hat{h}^2 + V^2}), \qquad \mu = x_r - \hat{h}\hat{\sigma}, \tag{5.2}$$

where

$$\bar{x}_r = \frac{1}{r} \sum_{i=1}^{r} x_i, \qquad V^2 = 4\left[1 + \frac{\sum_{i=1}^{r-1} (x_i - \bar{x}_r)^2}{r(x_r - \bar{x}_r)^2} \right],$$

and $\hat{h}$ is the solution of

$$\frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}\hat{h}^2\right)}{\frac{1}{\sqrt{2\pi}} \int_{\hat{h}}^{\infty} \exp\left(-\tfrac{1}{2}z^2\right) dz} = \frac{-r}{n-r}\hat{h} + \frac{2r}{n-r} \frac{\hat{h} + \sqrt{\hat{h}^2 + V^2}}{V^2}. \tag{5.2.1}$$

It is easy to show that the right and left sides of (5.2.1) are monotone decreasing and increasing respectively. This implies the uniqueness of the solution and also affords a simple method of solving (5.2.1) with the aid of a table of ordinates and areas of the standardized normal distribution. Despite the fairly formidable appearance of (5.2.1) the solution goes quickly.

It is also simple and interesting to calculate the asymptotic efficiency of the estimate of μ from a truncated sample when σ is known (the efficiency being considered relative to a completely known sample). For selected values of $q = r/n$, approximate efficiencies are given in thet following table. Approximate efficiency of $x_{[qn]}$, the sample percentage point, in estimating μ is also given to indicate the extent to which one gains by using the other $(r - 1)$ observations in the estimation procedure.

q	.1	.2	.3	.4	.5	.6	.7	.8	.9
Eff $\hat{\mu}$	.36	.53	.66	.75	.82	.88	.91	.95	.98
Eff $x_{[qn]}$	.33	.49	.57	.62	.64	.62	.57	.49	.33

Acknowledgment. The author is indebted to Dr. A. M. Mood who suggested this problem.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] H. Cramér, "A contribution to the theory of statistical estimation," *Skandinavisk Aktuarietidskrift*, Vol. 29 (1946), pp. 85–94.

[3] A. Hald, "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point," *Skandinavisk Aktuarietidskrift*, Vol. 32 (1949), pp. 119–34.

[4] A. M. Mood, "An asymptotic distribution for a mortality problem," *Rand Corporation document RM-111*, 1949, unpublished.

[5] A. C. Cohen, "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 557–569.

ON THE COMPARISON OF SEVERAL EXPERIMENTAL CATEGORIES WITH A CONTROL[1]

By Edward Paulson

University of Washington

Summary. This paper investigates certain statistical problems arising in the determination of the "best" of k categories when comparing $k - 1$ experimental categories with a standard or control. The discussion is limited to the case of a single stage sampling procedure with an equal number of observations on each of the k categories. Results both of an exact and of an approximate nature are obtained when (a) the observations with each category are normally distributed, and (b) the observations with each category have a binomial distribution.

1. Introduction. In this paper we will be concerned with the problem of the selection of one of the k categories $\Pi_1, \Pi_2, \cdots, \Pi_k$ as best when category Π_1 plays a special role, since it represents the standard or control, while $\Pi_2, \Pi_3, \cdots, \Pi_k$ represent $k - 1$ experimental categories. For the type of application we have in mind, the k categories might represent k varieties of wheat, or k drugs, or k machines; the "goodness" of a category will depend on some parameter of the probability distribution associated with that category. The experimental categories can be classified into two groups: one group consisting of those categories which are superior to Π_1 and a second group consisting of those experimental categories which are inferior to or at most equal to Π_1. In such a situation it will usually be desirable to have special protection against the selection of an experimental category as best when it actually is inferior to Π_1. This will be accomplished by requiring that the statistical procedure used provide a special assurance that Π_1 will be selected as best if the second group happens to consist of all $k - 1$ experimental categories, that is, none of the experimental categories is superior to Π_1. Situations of this type are believed to be fairly common in experiments in medicine and agriculture.

We will therefore consider the following statistical problem: given a sample consisting of kn independent observations $\{x_{ij}\}$ $(i = 1, 2, \cdots, k; j = 1, 2, \cdots, n)$, where x_{ij} is the jth observation with category Π_i, to devise a statistical procedure for selecting one out of the k categories as best so that if none of the experimental categories $\Pi_2, \Pi_3, \cdots, \Pi_k$ is actually "superior" to Π_1, then the probability that Π_1 is selected will be $\geqq 1 - \alpha$. We will also consider the related problem of deciding how large a sample will be required so that when one of the experimental categories is really superior to all the others including Π_1 by a specified amount the probability will also be $\geqq 1 - \beta$ that this experimental category will be selected as best. The constants α and β might be considered as

[1] Work on this paper done under the sponsorship of the office of Naval Research.

roughly analogous to the type I and type II errors in the Neyman-Pearson theory of testing a hypothesis. However the present problem is of a multiple decision type, and is not equivalent to one involving the testing of a hypothesis unless $k = 2$.

For the most part the discussion will be confined to the normal case, when the n observations with category Π_i $(i = 1, 2, \cdots, k)$ are assumed to be normally and independently distributed with mean m_i and common variance σ^2; the best category is defined to be the one associated with the greatest value of m_i. A brief discussion will also be given of the binomial case, when each observation with category Π_i is classified as a "success" or "failure" with a probability P_i of being a "success"; the best category is defined to be the one associated with the greatest value of P_i.

2. The normal case with known variance. We will first treat the problem when σ is assumed to be known a priori. Let $\bar{x}_i = \sum_{j=1}^{n} x_{ij}/n$, $\bar{x}^* = \max(\bar{x}_2, \bar{x}_3, \cdots, \bar{x}_k)$, $\alpha_1 = \alpha/(k-1)$, and for any a $(0 < a < 1)$ let ν_a be defined by the equation

$$\frac{1}{\sqrt{2\pi}} \int_{\nu_a}^{\infty} e^{-\frac{1}{2}t^2}\, dt = a.$$

Let Π^* be the experimental category whose mean is $\bar{x}^*$, and let λ be a constant whose value will be determined in a moment. The following statistical procedure is proposed for the selection of the best category:

$$\text{(1)}\qquad \begin{aligned} &\text{If } \bar{x}^* - \bar{x}_1 \geqq \lambda\sigma\sqrt{\frac{2}{n}}, \text{ select } \Pi^*;\\ &\text{If } \bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{\frac{2}{n}}, \text{ select } \Pi_1. \end{aligned}$$

We now complete the specification of the statistical procedure by determining λ. It is obvious that when $m_1 \geqq \max(m_2, m_3, \cdots, m_k)$, the greatest lower bound of the probability that $\bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{2/n}$ will occur when $m_1 = m_2 = \cdots = m_k$. In order to evaluate $P\{\bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{2/n} \mid m_1 = m_2 = \cdots = m_k\}$ we use the fact that $\bar{x}^*$ and $\bar{x}_1$ are independent, and find, after some simplifications,

$$\begin{aligned} P\left\{\bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{\frac{2}{n}} \,\middle|\, m_1 = m_2 = \cdots = m_k\right\} \\ &= \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}r^2}}{\sqrt{2\pi}} \int_{-\infty}^{r+\lambda\sqrt{2}} (k-1)\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \left[\int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt\right]^{k-2} dz\, dr \\ &= \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}r^2}}{\sqrt{2\pi}} \left[\int_{-\infty}^{r+\lambda\sqrt{2}} \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}}\, dt\right]^{k-1} dr. \end{aligned}$$

The constant λ will therefore be given as the root of the equation

$$\text{(2)}\qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}r^2} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r+\lambda\sqrt{2}} e^{-\frac{1}{2}t^2}\, dt\right]^{k-1} dr = 1 - \alpha.$$

From equation (2) it is possible to tabulate the values of λ as a function of α and k. Pending the construction of adequate tables, we will give an approximate method for finding λ with a definite limit of error. For this purpose, let A_i stand for the event $\bar{x}_i - \bar{x}_1 \geqq \lambda\sigma\sqrt{2/n}$. We have $P\{\bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{2/n}\} = 1 - P(A_2 + A_3 + \cdots + A_k)$, where the probabilities involved are to be calculated for the case when all the k means are equal. Making use of Bonferroni's Inequality (see [1], p. 75) we have

$$1 - \sum_{i=2}^{k} P(A_i) + \sum_{i}\sum_{\substack{j \\ j \geq i}} P(A_i \cdot A_j) \geqq P\left\{\bar{x}^* - \bar{x}_1 \leqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} \geqq 1 - \sum_{i=2}^{k} P(A_i).$$

Due to the symmetry when the means are equal, this becomes

$$\begin{aligned} &1 - (k-1)P(A_2) + \frac{(k-1)(k-2)}{2} P(A_2 \cdot A_3) \\ &\qquad \geqq P\left\{\bar{x}^* - \bar{x}_1 \leqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} \geqq 1 - (k-1)P(A_2). \end{aligned} \tag{3}$$

Since $(\bar{x}_2 - \bar{x}_1)$ and $(\bar{x}_3 - \bar{x}_1)$ have a bivariate normal distribution with correlation $= \frac{1}{2}$, we obtain

$$P(A_2) = \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-\frac{1}{2}t^2} dt,$$

$$P(A_2 A_3) = \frac{1}{\pi\sqrt{3}} \int_{\lambda}^{\infty} \int_{\lambda}^{\infty} e^{-\frac{2}{3}(x^2 - xy + y^2)}\, dx\, dy.$$

If we use as the approximate value for λ the solution $\tilde{\lambda}$ of the equation

$$\frac{1}{\sqrt{2\pi}} \int_{\tilde{\lambda}}^{\infty} e^{-\frac{1}{2}t^2} dt = \alpha_1, \tag{4}$$

then from (3) it follows that $P(\tilde{\lambda}) = P\{\bar{x}^* - \bar{x}_1 < \tilde{\lambda}\sigma\sqrt{2/n}\}$ will exceed $1 - \alpha$ by an amount which is not greater than $\frac{1}{2}(k-1)(k-2)P(A_2 A_3)$. This quantity can be calculated from the tables of the volumes of the normal bivariate surface [2]. The calculations for several values of α and k are summarized in Table I. It appears that the approximation yields good results for values of α which ordinarily are of interest if k is not too large, say $\leqq 6$.

Any statistical procedure for selecting one of the k categories can, of course, lead to other types of error than that of selecting an experimental category as the best when it actually is inferior to the standard or control. In particular, the error in not selecting a particular experimental category as best when it actually is superior to all the other experimental categories and the standard or control by at least a specified amount is of considerable interest. For a fixed value of α, this new type of error can be reduced only by increasing n, the sample size. Suppose for convenience that Π_k is the particular experimental category that exceeds the others by an amount Δ; that is, $m_k \geqq \max(m_1, m_2, \cdots, m_{k-1}) + \Delta$. Using the statistical procedure of (1), it is easy to see that for a

fixed λ, Δ, k, and n the greatest lower bound of the probability that Π_k will be selected as best will occur when $m_1 = m_2 = \cdots = m_{k-1} = m$ and $m_k = m + \Delta$.

TABLE I

Limits for $P(\bar{\lambda})$

$k \backslash \alpha$	.02	.05
3	$\bar{\lambda} = 2.326$ $.981 \geqq P(\bar{\lambda}) \geqq .98$	$\bar{\lambda} = 1.960$ $.955 \geqq P(\bar{\lambda}) \geqq .95$
6	$\bar{\lambda} = 2.652$ $.984 \geqq P(\bar{\lambda}) \geqq .98$	$\bar{\lambda} = 2.326$ $.963 \geqq P(\bar{\lambda}) \geqq .95$

If we denote this greatest lower bound by $P(n; \lambda, k, \Delta)$, we easily obtain

$$
\begin{aligned}
&P(n; \lambda, k, \Delta) \\
&= P\left\{\bar{x}_k - \bar{x}_1 > \lambda\sigma\sqrt{\frac{2}{n}} \text{ and } \bar{x}_k > \max(\bar{x}_2, \bar{x}_3, \cdots, \bar{x}_{k-1}) \mid m_k = m_1 + \Delta\right\} \\
(5)\quad &= \int_{-\infty}^{\infty}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w-\lambda\sqrt{2}} e^{-\frac{1}{2}t^2}\,dt\right] \\
&\qquad \left\{\int_{-\infty}^{\infty}(k-2)\frac{e^{-\frac{1}{2}s^2}}{\sqrt{2\pi}}\left[\int_{-\infty}^{s}\frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}}\,dt\right]^{k-3} ds\right\}\frac{e^{-\frac{1}{2}(w-(\Delta/\sigma)\sqrt{n})^2}}{\sqrt{2\pi}}\,dw \\
&= \int_{-\infty}^{\infty}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w+(\Delta/\sigma)\sqrt{n}-\lambda\sqrt{2}} e^{-\frac{1}{2}t^2}\,dt\right]\cdot\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w+(\Delta/\sigma)\sqrt{n}} e^{-\frac{1}{2}t^2}\,dt\right]^{k-2}\frac{e^{-\frac{1}{2}w^2}}{\sqrt{2\pi}}\,dw.
\end{aligned}
$$

In order to decide in advance of the experiment how large a sample should be taken, we might try to find n so that $P(n; \lambda, k, \Delta) = 1 - \beta$ where α, β, and Δ are determined by practical considerations depending on the particular experimental situation, and λ is found from (2) or (4). It will be very difficult to find n directly from (5) until tables are made available. However we can usually obtain a good approximation $\tilde{n}$ to the required value by solving the equation

$$
(6)\quad P\left\{\bar{x}_k - \bar{x}_1 \leqq \lambda\sigma\sqrt{\frac{2}{n}}\,\middle|\, m_k = m_1 + \Delta\right\} = \frac{1}{\sqrt{2\pi}}\int_{(\Delta/\sigma)\sqrt{n/2}-\lambda}^{\infty} e^{-\frac{1}{2}t^2}\,dt = \beta.
$$

The solution can be written $\tilde{n} = (2\sigma^2/\Delta^2)(\lambda + \nu_\beta)^2$ which reduces to $\tilde{n} = (2\sigma^2/\Delta^2)(\nu_{\alpha_1} + \nu_\beta)^2$ when $\bar{\lambda}$ is used for λ. The adequacy of this approximation can be estimated with the help of the inequality[2]

$$
\begin{aligned}
&P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} - (k-2)P\{\bar{x}_k < \bar{x}_2\} \\
(7)\quad &\qquad \leqq P\{n; \lambda, k, \Delta\} \leqq P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda\sigma\sqrt{\frac{2}{n}}\right\},
\end{aligned}
$$

[2] The writer is indebted to the referee for this inequality, which is an improvement over the one originally used.

which holds when $m_1 = m_2 = \cdots = m_{k-1}$, $m_k = m_1 + \Delta$. To derive this inequality, it is obvious from the definition of $P\{n; \lambda, k, \Delta\}$ that $P\{n; \lambda, k, \Delta\} \leqq P\{\bar{x}_k - \bar{x}_1 > \lambda\sigma\sqrt{2/n}\}$, while from Bonferroni's Inequality we have

$$P\{n; \lambda, k, \Delta\} \geqq 1 - P\left\{\bar{x}_k - \bar{x}_1 \leqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} - \sum_{j=2}^{k-1} P\{\bar{x}_k \leqq \bar{x}_j\}$$

$$= P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} - (k-2)P\{\bar{x}_k \leqq \bar{x}_2\}.$$

Hence when $\tilde{n}$ is found from (6), it follows from (7) that $[1 - P(\tilde{n}; \lambda, k, \Delta)]$ will exceed β by an amount which is less than

$$(k-2)P\{\bar{x}_k \leqq \bar{x}_2\} = \frac{(k-2)}{\sqrt{2\pi}} \int_{(\Delta/\sigma)\sqrt{\tilde{n}/2}}^{\infty} e^{-\frac{1}{2}t^2}\,dt.$$

We have attempted to indicate the adequacy of the approximation $\tilde{n}$ found from (6) by computing the upper bound for $[1 - P(\tilde{n}; \lambda, k, \Delta)]$ for several

TABLE II

Upper bound for $[1 - P(\tilde{n}; \tilde{\lambda}, k, \Delta)]$

k	α / β	.05	.02
3	.20	.2025	.2008
	.05	.0502	.0500
6	.20	.2031	.2010
	.05	.0501	.0500

values of α, β and k; in these calculations the value of $\tilde{\lambda}$ given in Table I was used for λ. The calculations are summarized in Table II.

It appears that for $\beta \leqq .20$, $\alpha \leqq .05$, the upper bound for $[1 - P(\tilde{n}; \lambda, k, \Delta)]$ exceeds the corresponding β by a small amount which can be neglected for most practical purposes.

3. The normal case with unknown variance. The case when σ is unknown will now be briefly considered. Let

$$s^2 = \sum_{i=1}^{k} \sum_{\alpha=1}^{n} \frac{(x_{i\alpha} - \bar{x}_i)^2}{k(n-1)}$$

be the pooled estimate of σ^2 based on $n' = k(n-1)$ degrees of freedom. The statistical procedure in (1) is modified as follows:

$$\text{select } \Pi^* \text{ if } \bar{x}^* - \bar{x}_1 \geqq \lambda_n s\sqrt{\frac{2}{n}};$$

(8)

$$\text{select } \Pi_1 \text{ if } \bar{x}^* - \bar{x}_1 < \lambda_n s\sqrt{\frac{2}{n}}.$$

The exact equation that λ_n must satisfy in order to have $P\{\bar{x}^* - \bar{x}_1 < \lambda_n s \sqrt{2/n} \mid m_1 = m_2 = \cdots = m_k\} = 1 - \alpha$ and an explicit expression for $P(n; \lambda_n, k, \Delta/\sigma)$, the probability that Π_k will be selected as best when $m_1 = m_2 = \cdots = m_{k-1}$, $m_k/\sigma = m_1/\sigma + \Delta/\sigma$, can easily be found by a procedure similar to that used for (2) and (5); however the results are complicated and instead we will proceed directly to discuss approximate procedures. Let C_j denote the event $\bar{x}_j - \bar{x}_1 > \lambda_n s \sqrt{2/n}$ $(j = 2, 3, \cdots, k)$, and let $t_{n'}$ denote a random variable having the t distribution with n' degrees of freedom. With the aid of tables of the t distribution an approximation $\tilde{\lambda}_n$ to λ_n can be found so that $P\{\bar{x}_2 - \bar{x}_1 > \tilde{\lambda}_n s \sqrt{2/n}\} = P\{t_{n'} > \tilde{\lambda}_n\} = \alpha/(k-1) = \alpha_1$. Then the probability that Π_1 will be selected as best when all the means are equal will exceed $1 - \alpha$ by an amount which is less than $\frac{1}{2}(k-1)(k-2)P(C_2 \cdot C_3)$. For bounds on the second type of error, we have

$$P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda_n s \sqrt{\frac{2}{n}} \,\middle|\, \frac{m_k}{\sigma} = \frac{m_1}{\sigma} + \frac{\Delta}{\sigma}\right\} - (k-2)P\left\{\bar{x}_k \leqq \bar{x}_2 \,\middle|\, \frac{m_k}{\sigma} = \frac{m_1}{\sigma} + \frac{\Delta}{\sigma}\right\}$$

$$\leqq P\left\{n; \lambda_n, k, \frac{\Delta}{\sigma}\right\} \leqq P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda_n s \sqrt{\frac{2}{n}} \,\middle|\, \frac{m_k}{\sigma} = \frac{m_1}{\sigma} + \frac{\Delta}{\sigma}\right\}.$$

All these inequalities are easily obtained as in Section 2. To evaluate the bound for the first type of error, a good approximation can usually be found by regarding s/σ to be normally distributed with mean 1 and variance $1/(2n')$; using this approximation it is easy to verify that

$$P(C_2 \cdot C_3) = P\left\{U \geqq \lambda_n \sqrt{\frac{2n'}{2n' + \lambda_n^2}} \quad \text{and} \quad V \geqq \lambda_n \sqrt{\frac{2n'}{2n' + \lambda_n^2}}\right\},$$

where (U, V) has a bivariate normal distribution with zero means, unit variances and correlation $\rho = (n' + \lambda_n^2)/(2n' + \lambda_n^2)$. This same device might also be used to approximate the upper and lower bounds for $P(n; \lambda_n, k, \Delta/\sigma)$ as an alternative to evaluating the bounds by using tables of the non-central t distribution. Finally, to obtain the value of n so that $P(n; \tilde{\lambda}_n, k, \Delta/\sigma) = 1 - \beta$ a good first approximation will usually be given by $n_0 = (2\sigma^2/\Delta^2)(\nu_{\alpha_1} + \nu_\beta)^2$; after computing $\tilde{\lambda}_{(n_0)}$ and the corresponding upper and lower bounds for $P(n_0; \tilde{\lambda}_{(n_0)}, k, \Delta/\sigma)$, the first approximation n_0 can be modified if necessary and the process iterated.

It should be noted that in order to find the sample size required to control the second type of error, either an approximate value of σ must be known from past experience, or else it must be sufficient for the practical problem under consideration to know the probability of selecting the best experimental category as a function of the ratio of Δ to σ. It is possible to eliminate the dependence of the result on σ by making use of a two-stage sampling scheme due to Stein [3]; this and other sequential procedures may be considered in another paper.

4. The binomial case. In this section a brief treatment of the binomial case will be given, based on the use of the inverse sine transformation. That is, we

will use the fact that if $\bar{p}$ is the observed proportion of successes in n independent trials with a constant probability P of a success, then arc sin $\sqrt{\bar{p}}$ is for large n approximately normally distributed with mean (arc sin $\sqrt{P}$) and variance $1/(4n)$ (provided the angle is given in radian measure). This transformation was previously used by W. Allen Wallis and the present writer [4] to design experiments for comparing the percentages associated with one experimental and one standard category. The material in this section can be regarded as one possible extension of that work to the case where we are dealing with more than one experimental category.

Let r_i be the number of successes in the n observations with category Π_i. Let $\bar{p}_i = (r_i/n)$, let $u_i = \text{arc sin} \sqrt{\bar{p}_i}$ and let P_i = the true probability of a success with category Π_i. Let $\bar{p}^* = \max (\bar{p}_2, \bar{p}_3, \cdots, \bar{p}_k)$, $u^* = \max (u_2, u_3, \cdots, u_k)$, and let Π^* be the experimental category with observed percentage of successes $\bar{p}^*$. If there should happen to be more than one category with the observed percentage of successes $= \bar{p}^*$, select Π^* at random from the subset having $\bar{p}_i = \bar{p}^*$.

We now propose the following statistical procedure for selecting one of the k categories.

$$
(9)\qquad
\begin{aligned}
&\text{select } \Pi^* \text{ if } u^* - u_1 \geqq \lambda \sqrt{\frac{1}{2n}};\\
&\text{select } \Pi_1 \text{ if } u^* - u_1 < \lambda \sqrt{\frac{1}{2n}},
\end{aligned}
$$

where λ is to be chosen so that if $P_1 \geqq \max (P_2, P_3, \cdots, P_k)$ the probability that Π_1 is selected as best will be $\geqq 1 - \alpha$. We assume that n is large enough so that the set $\{u_i\}$ can be regarded as normally distributed with common variance $1/(4n)$ and means arc sin $\sqrt{P_i}$. Therefore the problem is once again essentially equivalent to the normal case with known variance, which was treated in Section 2, and the value of λ is given by the solution of (2). To find the sample size n so that if $P_1 = P_2 = \cdots = P_{k-1} = P$ and $P_k = P + \delta$ $(\delta > 0)$, the probability that Π_k is selected as best will equal $1 - \beta$, set $\Delta = \text{arc sin} \sqrt{P + \delta} - \text{arc sin} \sqrt{P}$, and the required n will be given by (5) when $\sigma\sqrt{2/n}$ is replaced by $\sqrt{1/2n}$.

For values of α, β, and k so that $\alpha \leqq .05$, $\beta \leqq .20$, and $k \leqq 6$, it has been shown in Section 2 that if we use approximate values of λ and n given by (4) and (6), the change in the probabilities considered will be small, and will ordinarily be of little practical importance. Using the notation of the last section, (4) and (6) for the binomial case are equivalent to $\lambda = \nu_{\alpha_1}$ and $n = (\nu_{\alpha_1} + \nu_\beta)^2/(2\Delta^2)$.

We conclude this section by discussing a specific problem. Consider a situation in which we are interested in investigating the effect of three experimental treatments on a certain disease, where it is known from previous experience that the probability of survival with the standard treatment is of the order of magnitude of .75. The problem we wish to consider is that of designing a statistical procedure (based on a single stage of sampling) for selecting one of the 4 treatments which will have the following properties: (a) the probability of

selecting an experimental treatment as best when in fact it is inferior to the standard treatment is to be $\leqq .05$; (b) if one of the experimental treatments should happen to increase the probability of survival to .90, while the probabilities of survival for the three other treatments is $\leqq .75$, then the probability that the superior experimental treatment will be selected as best should be $\geqq .95$. Upon setting $\alpha = .05$, $\beta = .05$, and $k = 4$ we find $\lambda = 2.128$, $\Delta = .202$, $n = (2.128 + 1.645)^2/(2\Delta^2) = 174$.

The required statistical procedure having properties (a) and (b) is the following. A group of 696 animals are all innoculated with the specific disease under consideration, and then the animals are subdivided in some random manner into 4 groups each consisting of 174 animals. The first of these groups is given the standard treatment, and the remaining groups each receive one of the experimental treatments. After the experiment is completed, if $\text{arc sin } \sqrt{\bar{p}^*} - \text{arc sin } \sqrt{\bar{p}_1} \geqq 2.128/\sqrt{2(174)} = .114$, we conclude that the experimental treatment with observed percentage of success $= \bar{p}^*$ is best, otherwise we conclude that the standard treatment is really better than any of the experimental treatments.

REFERENCES

[1] William Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, 1950.

[2] K. Pearson, *Tables for Statisticians and Biometricians*, Part II, 1st ed., Cambridge University Press, 1931.

[3] Charles Stein, "A two sample test for a linear hypothesis whose power is independent of the variance," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 253–258.

[4] E. Paulson and W. A. Wallis, "Planning and analyzing experiments involving two percentages," *Selected Techniques of Statistical Analysis*, McGraw-Hill Book Co., pp. 247–265.

ON THE MOST ECONOMICAL SAMPLE SIZE FOR CONTROLLING THE MEAN OF A POPULATION

By H. Weiler

New South Wales University of Technology, Sydney, Australia

Summary. For quality control charts controlling the mean of a population either small samples may be taken out at frequent intervals or larger samples at less frequent intervals. In this paper, a simple formula is derived by which the most suitable sample size can be determined, leading to the detection of any given change of the population mean with a minimum of inspection.

1. Introduction. Consider a normal variate x representing some measure of a mass-produced article, and suppose that control limits similar to those used in the preparation of control charts [1], [2] are to be determined to detect changes of the population mean of x. Such control limits may be placed on a chart similar to that used in control chart analysis.

After the mean and standard deviation of a population have been estimated by means of an initial large sample, smaller samples of fixed size N are taken during the production, and their arithmetic means $\bar{x} = \sum x/N$ are calculated. A chart is then constructed with control limits $m \pm 3\sigma/\sqrt{N}$, where m and σ are the estimates of the population mean and S.D. obtained from the original large sample. The various values of $\bar{x}$ are then entered in the chart in chronological order, and as soon as one such value falls outside the control limits, production is stopped to allow investigation.

The aim of this paper is to determine the most economical sample size, that is, that value of N which would indicate a change of the population mean after a minimum amount of inspection. It will be found that the most economical sample size depends on the amount by which the population mean has changed. Thus, if the population mean changes from m to $m + k\sigma$, while σ remains constant, the most economical sample size $N = n$ will be a function of k. In particular, it will be shown that this function is

$$n = \frac{12.0}{k^2} \text{ when the control limits are } m \pm \frac{3.09\sigma}{\sqrt{n}}$$

$$n = \frac{11.1}{k^2} \text{ when the control limits are } m \pm \frac{3\sigma}{\sqrt{n}}$$

$$n = \frac{6.65}{k^2} \text{ when the control limits are } m \pm \frac{2.58\sigma}{\sqrt{n}}$$

$$n = \frac{4.4}{k^2} \text{ when the control limits are } m \pm \frac{2.33\sigma}{\sqrt{n}}.$$

Tables are calculated, giving the value of n for other control limits and giving also the average amount of inspection in each case.

Finally, a new chart, based on two sets of control limits, is discussed briefly.

2. The average amount of inspection for a given N. Let m be the original mean and σ the S.D. of the population, and let $m \pm B\sigma/\sqrt{N}$ be the control limits adopted for the arithmetic mean $\bar{x} = \sum x/N$ of a sample of given size N.

TABLE 1

N	$3.09 - 0.4\sqrt{N}$	P	S(N)	A(N)
1	2.690	0.0036	278	278
2	2.524	0.0058	173	345
3	2.396	0.0083	121	362
4	2.290	0.0110	91	364 Max
5	2.195	0.0141	71	355
9	1.890	0.0294	34	306
16	1.490	0.0681	14.7	235
25	1.090	0.1379	7.26	182
36	0.690	0.2451	4.08	147
49	0.290	0.3859	2.59	127
64	−0.110	0.5438	1.84	118
75	−0.375	0.6460	1.55	116 Min
81	−0.51	0.6950	1.44	117

If the population mean changes from $\mu = m$ to $\mu = m + k\sigma$ ($k > 0$), the probability that $\bar{x}$ exceeds the upper control limit $m + B\sigma/\sqrt{N}$ is (assuming that σ remains unchanged)

$$P = P\left(\bar{x} \geqq m + \frac{B\sigma}{\sqrt{N}} \middle| \mu = m + k\sigma\right)$$

$$= P\left(\bar{x} - m - k\sigma \geqq \frac{B\sigma}{\sqrt{N}} - k\sigma \middle| \mu = m + k\sigma\right) \tag{1}$$

$$= P\left(\frac{\bar{x} - m - k\sigma}{\sigma/\sqrt{N}} \geqq B - k\sqrt{N} \middle| \mu = m + k\sigma\right) = P(z \geqq B - k\sqrt{N}),$$

where z is the standardized normal variate (mean zero and S.D. one).

Thus, when μ becomes equal to $m + k\sigma$, about $100P$ samples in every 100 samples, or one in every $1/P$ samples will give a mean $\bar{x}$ above the upper control limit. It follows that on the average $S(N) = 1/P$ samples, or $A(N) = N/P$ articles have to be tested before a change of the mean from m to $m + k\sigma$ can be expected to be revealed.

3. Example. For illustration, take the example $k = 0.4$, $B = 3.09$. We obtain Table 1 for various values of N (using normal probability tables).

The example shows the following interesting points:

(a). *Suppose that the population mean changes by* $+0.4$ *standard deviations.* If the chart is based on the customary sample size used for control charts [3], namely, $N = 4$ or $N = 5$, about 360 articles must be tested before detection of the change can be expected. On the other hand, if the control chart is based on a sample size between 50 and 80 (say), about 120 items only are required to indicate the change. The usual engineering practice of using charts for small samples requires thus about three times as much inspection as would be required with a chart for a suitably large sample size.

(b). *Suppose that the population mean does not change.* In that case $\bar{x}$ will fall above the upper control limit about once in 1000 samples. This means that with sample size 4 a "false alarm" will be raised about once for every 4000 articles tested. With a sample size 75, on the other hand, a "false alarm" will be raised only once for about 75,000 articles tested.

The two points raised suggest that in certain cases it may be of advantage to deviate from the usual practice of using small sample control charts. A third argument in favor of large samples is that the control limits are based on the assumption that $\bar{x}$ is normally distributed. This assumption is usually satisfied with great accuracy when the sample is large, but may not be justified when the sample is small.

The above arguments hold also for other values of k and B, and we may state that, unless special reasons exist for making the samples small, the sample size N should be chosen such that the average amount of inspection $A(N)$ becomes a minimum.

4. The minimum amount of inspection. We define the most economical sample size n as that value of N for which the average amount of inspection $A(N)$ required to detect a given change of the population mean becomes a minimum.

If the standardized normal probability density is denoted by $\varphi(z) = e^{-\frac{1}{2}z^2}/\sqrt{2\pi}$, we have

$$A(N) = N/P = N \Big/ \int_{B-k\sqrt{N}}^{\infty} \varphi(z)\,dz. \tag{2}$$

Differentiating with regard to N, we have

$$\frac{dA}{dN} = \frac{\int_{B-k\sqrt{N}}^{\infty} \varphi(z)\,dz - \frac{1}{2}k\sqrt{N}\,\varphi(B - k\sqrt{N})}{\left[\int_{B-k\sqrt{N}}^{\infty} \varphi(z)\,dz\right]^2}. \tag{3}$$

The condition for $A(N)$ to be a minimum is $(dA/dN)_{N=n} = 0$, which reduces to

$$P(u) = \int_u^{\infty} \varphi(z)\,dz = \tfrac{1}{2}(B - u)\varphi(u), \tag{4}$$

where $u = B - k\sqrt{n}$.

Equation (4) is easily solved, using tables of ordinates and integrals of the normal distribution. To do this, we put the minimum condition (4) in the form

$$Q = \frac{2P(u)}{\varphi(u)} = B - u. \tag{5}$$

The left side of this equation can then be calculated for any value of u and the values n, $S(n)$, and $A(n)$, can be deduced. We have

$$n = \frac{(B-u)^2}{k^2} = \frac{Q^2}{k^2}, \tag{6}$$

$$S(n) = \frac{1}{P}, \tag{7}$$

$$A(n) = \frac{n}{P} = \frac{Q^2}{Pk^2}. \tag{8}$$

The values of $S(n)$, $k^2A(n)$, and k^2n are shown in Table 2 for various values of u.

It can be seen from this table that for every value $B > 2.24$ two values N exist which satisfy the condition $dA/dN = 0$. Only the larger value of N, however, corresponds to a minimum amount of inspection. (The smaller value of N corresponds to a maximum of $A(N)$). Thus, for $B = 3$ we find $n_1 = 11.1/k^2$ and $n_2 = 0.606/k^2$. The amounts of inspection corresponding to n_1 and n_2 are $A(n_1) = 17.65/k^2$ and $A(n_2) = 46/k^2$ respectively, which shows that samples of size n_2 would lead to a much larger amount of inspection than samples of size n_1. The lower part of Table 2, corresponding to values of u greater than 0.6, can therefore be ignored.

Besides this, we notice that no values of u exist for which B is less than 2.24. This means that for such values of B no value of N exists which would make dA/dN equal to zero. This case will be discussed later (Section 6).

5. Discussion of the special case B = 3.09. The values $B = 3.09$, $B = 3$, $B = 2.58$, $B = 2.33$, are of special interest because they correspond to the most frequently used control limits. In particular, we have for $B = 3.09$: $n = 12.0/k^2$, $S(n) = 1.55$, $A(n) = 18.6/k^2$. The values of n and $A(n)$ for various values of k are tabulated below.

k	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8
n	300	133	75	54	37	28	21	17	12	8	6	5	4
$A(n)$	464	206	116	74	52	38	29	23	19	13	9.5	7.3	5.7

Alternatively, we may plot n and $A(n)$ as functions of k; they become straight lines when plotted on log log paper.

When the most economical sample size is taken, the average number of samples required for the detection of a change in the population mean is the same for all values of k; it is equal to $S(n) = 1.55$ when $B = 3.09$. This form of

TABLE 2

u	$P = \int_u^\infty \varphi(z)dz$	$\varphi(u)$	$Q = \frac{2P}{\varphi}$	$B = Q + u$	$nk^2 = Q^2$	$S(n) = \frac{1}{P}$	$A(n)k^2 = \frac{Q^2}{P}$
−0.5	0.692	0.352	3.92	3.42	15.4	1.45	22.3
−0.4	0.655	0.368	3.56	3.16	12.6	1.53	19.2
−0.37	0.6443	0.3725	3.46	3.09	12.0	1.55	18.6
−0.33	0.6293	0.3778	3.33	3.00	11.1	1.59	17.65
−0.3	0.618	0.381	3.24	2.94	10.3	1.62	17.0
−0.2	0.579	0.391	2.96	2.76	8.8	1.73	15.2
−0.1	0.540	0.397	2.72	2.62	7.4	1.85	13.7
−0.07	0.5279	0.3980	2.655	2.585	7.04	1.89	12.2
0.0	0.500	0.399	2.51	2.51	6.3	2.00	12.6
0.1	0.460	0.397	2.32	2.42	5.4	2.17	11.7
0.2	0.421	0.391	2.15	2.35	4.6	2.38	10.9
0.24	0.4052	0.3876	2.095	2.335	4.36	2.47	10.75
0.3	0.382	0.381	2.00	2.30	4.0	2.62	10.5
0.4	0.345	0.368	1.88	2.28	3.5	2.90	10.1
0.5	0.308	0.352	1.75	2.25	3.06	3.24	9.94
0.6	0.274	0.333	1.64	2.24	2.70	3.65	9.86
0.7	0.242	0.312	1.55	2.25	2.40	4.13	9.94
0.8	0.212	0.290	1.46	2.26	2.13	4.72	10.1
1.0	0.159	0.242	1.31	2.31	1.73	6.28	10.9
1.2	0.115	0.194	1.19	2.39	1.41	8.69	12.2
1.4	0.081	0.150	1.08	2.48	1.16	12.3	14.3
1.6	0.055	0.111	0.99	2.59	0.98	18.2	17.8
2.0	0.0228	0.054	0.844	2.844	0.71	43.8	31
2.22	0.0132	0.0339	0.78	3.00	0.606	76	46

control chart is thus very efficient, for it will indicate a change in about 2 out of 3 samples, whereas it will raise a false alarm (or type I error [4]) only in about one out of 1000 samples.

It appears from the above table that a control chart for small samples, say N between 4 and 10, is adequate only for the detection of changes of the mean greater than one standard deviation.

There is, of course, no need to adhere rigidly to the sample size given by the table, for in most cases the exact change (if any) of the population mean would not be known beforehand, but the table will give valuable information regarding the approximate size of the sample required.

Thus, in the case when $B = 3.09$, we would recommend the following sample sizes:

Change of Mean μ in S.D.'s	*Sample Size N*
0.2–0.3	100–300
0.3–0.4	70–150
0.4–0.5	50– 80
0.5–0.6	30– 60
0.6–0.7	25– 40
0.7–0.8	20– 30
0.8–1.0	10– 25

To detect changes larger than one standard deviation, any convenient size up to 10 could be taken.

6. The case when B < 2.24. When $B < 2.24$, the derivative dA/dN is different from zero for all values of N. This means that $A(N)$ has no relative minimum but increases with N for all values of N. The average amount of inspection is then smallest when $N = 1$, but, for reasons stated in Section 3, it is usually not advisable to take N smaller than, say, 4.

The only case of this type that might be of interest for the purpose of quality control is the case $B = 1.96$, because the control limits $m \pm 1.96\ \sigma/\sqrt{N}$ are often used as so-called inner limits [5]. The average amount of inspection is then (Section 2) $A(N) = N/P = N/P(z \geqq u)$, where $u = B - k\sqrt{N}$. This gives $N = (B - u)^2/k^2$ and $A(N) = (B - u)^2/(k^2P)$. Taking $B = 1.96$ and substituting different values for u, we obtain Table 3.

The table shows clearly that the average amount of inspection $A(N)$ decreases with the size N of the sample, that is, the smaller we make N the more economical will be the test.

7. A chart with two sets of control limits. When a chart with two sets of control limits is used, it is usually set up as follows.

After the mean m and S.D. σ of the population have been reliably estimated, a suitable sample size N is chosen, and inner limits ($m \pm 1.96\ \sigma/\sqrt{N}$) and outer limits ($m \pm 3.09\ \sigma/\sqrt{N}$) are calculated and entered in the chart. Production is stopped as soon as one $\bar{x}$ value falls outside the outer control limits. The main

purpose of the inner limits is to provide a first warning when a point falls outside these limits. Production is then not interrupted, but samples are taken more frequently in order to reach a decision without delay.

Now we have seen that small samples are most suitable for the inner limits while larger samples should be taken for the outer limits. It seems therefore indicated to construct a chart involving two sample sizes. (If a smaller sample size is used for the inner control limits, the terms "inner" and "outer" become misleading because the inner limits will then actually be wider than the outer limits.) To detect a change of the population mean from m to $m + k\sigma$, a chart may be prepared as follows. (1) Take samples of size $N = 4$ or 5 (or any other

TABLE 3
($B = 1.96$)

u	P	$B - u$	$Nk^2 = (B - u)$	$A(N)k^2$	$S(N) = \frac{1}{P}$
1.76	.0392	0.2	0.04	1.0	25.5
1.66	.0485	0.3	0.09	1.9	20.6
1.56	.0594	0.4	0.16	2.7	16.8
1.46	.0721	0.5	0.25	3.5	13.9
1.36	.0869	0.6	0.36	4.1	11.5
1.26	.104	0.7	0.49	4.7	9.6
1.16	.123	0.8	0.64	5.2	8.1
1.06	.145	0.9	0.81	5.6	6.9
0.96	.168	1.0	1.00	6.0	6.0
0.76	.224	1.2	1.44	6.4	4.5
0.56	.288	1.4	1.96	6.8	3.5
0.36	.359	1.6	2.56	7.1	2.8
0.16	.436	1.8	3.24	7.4	2.3
−0.04	.516	2.0	4.00	7.8	1.9
−0.54	.705	2.5	6.25	8.9	1.4

convenient small sample size) and construct a chart with control limits $m \pm 1.96\sigma/\sqrt{N}$. (2) Calculate a second set of control limits $m \pm 3.09\sigma/\sqrt{n}$, based on a sample size $n = \lambda N$, which is a multiple of N as close as possible to the most economical sample size $12.0/k^2$. (3) Calculate the means $\bar{x} = \sum x/N$ and $\bar{X} = \sum \bar{x}/\lambda$. (4a) If a value $\bar{X}$ falls outside the limits $m \pm 3.09\sigma/\sqrt{n}$, stop production and investigate; (4b) if a value $\bar{x}$ falls outside the limits $m \pm 1.96\sigma/\sqrt{N}$, do not interrupt production but take out samples frequently to reach a decision.

While it is true that the "inner" limits serve mainly to provide a first warning for a possible change of the population mean, they may be used also to reach a definite decision. If, for instance, two successive values of $\bar{x}$ fall above the upper inner limit, we may regard this as a significant indication for a change in the

population mean, because the probability for this to happen when the population mean is unchanged is only $(0.025)^2 = 0.0006$, which is less than the probability that a single value $\bar{X}$ falls above the upper outer limit $m + 3.09\sigma/\sqrt{n}$.

Again, it is easily shown that the probability that 4 out of 16 successive $\bar{X}$ values fall above the upper inner limit is about the same as the probability that a single $\bar{x}$ value falls above the upper outer limit. We are therefore justified in regarding such an occurrence as a significant indication for a change of the population mean.

8. Conclusion. When a single set of control limits is used for a chart controlling the mean of a population, the above theory leads to much larger samples than those usually taken in industry. However, even with a chart of this type, small samples are not always uneconomical, for there are other factors to be considered which are not covered in this paper.

For instance, it may be of advantage to divide samples into subgroups in order to detect changes due to definite anticipated causes, necessitating the use of smaller samples.

Again, if a change of the population mean from m to $m + k\sigma$ is anticipated and the sample size is determined accordingly, any unsuspected larger change would in the average be detected later than if a smaller sample size had been used.

On the other hand, if small changes of the population mean of a given order are anticipated and if it is unlikely that larger changes occur, the sample size should be calculated according to the above theory.

When it is convenient to use a more elaborate chart, containing two sets of control limits, the theory leads to the customary small samples for the one set ($m \pm 1.96\sigma/\sqrt{N}$) and to the above large samples for the other set ($m \pm 3.09\sigma/\sqrt{n}$). Any unexpected larger change of the population mean is then likely to be detected by means of the small samples.

REFERENCES

[1] W. A. SHEWHART, *The Economic Control of Quality of a Manufactured Product*, D. Van Nostrand, 1931.

[2] W. A. SHEWHART, *Contribution of Statistics to the Science of Engineering*, University of Pennsylvania Press, 1940.

[3] W. B. RICE, *Control Charts in Factory Management*, John Wiley and Sons, 1947.

[4] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses. I. Unbiassed critical regions of Type A and Type A_1," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1-37.

[5] B. P. DUDDING AND W. J. JENNETT, *Quality Control Charts*, No. 600R, British Standards Institution, 1942.

[6] AMERICAN SOCIETY FOR TESTING MATERIALS, *ASTM Manual on Quality Control of Materials*, Philadelphia, 1951.

[7] S. S. WILKS, "Determination of sample size for setting tolerance limits," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 91-96.

OPTIMUM ALLOCATION IN LINEAR REGRESSION THEORY

By G. Elfving

University of Helsingfors and Cornell University

Summary. If for the estimation of β_1, β_2 different observations ("sources") of form (1.1) are potentially available, each of them being repeatable as many times as we please, the question arises which of them the experimenter should utilize, and in what proportions. With appropriate optimality conventions the answer is the following. For the estimation of a single quantity of form $\theta = a_1\beta_1 + a_2\beta_2$ the optimum allocation comprises two sources only; for the estimation of both parameters, the corresponding number is two or three; the best proportions are indicated in Sections 2 and 4 below. Generalizations to more than two parameters and to observations at different costs are briefly discussed.

The problem is related to Hotelling's weighing problem [2] and to the topics treated by David and Neyman in [1].

1. Introduction. Consider an experimenter who wants to determine two unknown quantities β_1, β_2. We assume that for this purpose a certain number r of different *potential* observations are at his disposal, the outcomes of which are of form

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \eta_i \qquad (i = 1, \cdots, r); \tag{1.1}$$

here x_{i1}, x_{i2} denote known coefficients and η_i a random variable (the error term) with mean zero and standard deviation σ. (If the standard deviations of the different observations are proportional to known numbers $k_1, \cdots, k_r$, we have only to divide the equations (1.1) by these numbers in order to restore the situation of the text.) We assume, furthermore, that the experimenter may perform each of the observations as many times as he pleases, or not at all, all actual observations being uncorrelated. If he has decided upon a certain total n of actual observations, he is faced with the problem which of the potential ones should be performed, and in what number. As an application, the reader may think of a surveyor who wants to find the coordinates of a point by observing the direction to it from given surrounding points of known position; in this case the regression is, of course, only differentially linear.

In order to distinguish between the potential and the actual observations, we will in the following refer to the former as *sources* (of information), to the latter as *observations*. Since a source is essentially described by the coefficient vector $\mathbf{x}_i = (x_{i1}, x_{i2})$, we will also briefly speak of the source $\mathbf{x}_i$ or simply the source i. In the solution of any particular optimum allocation problem, those sources which are actually utilized will be called *relevant*, the others *irrelevant*.

The following normalization and idealization of our problem is mathematically convenient. Let the required number of observations on the ith source be

$n_i = np_i$; the p_i's are obviously multiples of $1/n$ fulfilling the conditions

$$p_i \geqq 0, \qquad \sum p_i = 1. \tag{1.2}$$

The mean of the observations on the ith source then has a regression equation which may be written

$$\bar{y}_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \frac{\eta_i}{\sqrt{p_i}} \qquad (i = 1, \cdots, r), \tag{1.3}$$

where η_i has variance σ^2/n. If a certain p_i is zero, the corresponding equation has to be left out of the system. For large n, the p_i's may be varied practically continuously over the range (1.2). Idealizing this feature we get a large-sample problem, which is essentially independent of σ and n. For simplicity we may finally assume $\sigma^2/n = 1$; in order to restore full generality we have only to reintroduce this factor in all variance and covariance formulas below. We are then faced with the following question:

Consider a planned set of observations of form (1.3) *with* $E(\eta_i) = 0$, $D^2(\eta_i) = 1$, *and with the weights* p_i *at our disposal, subject only to the conditions* (1.2). *What are the optimal* p_i*'s?*

The solution of this problem obviously presupposes a specification of the word "optimal," that is, a specification of the estimation problem at hand.

When applying the solution to practical problems, one has to remember that our large-sample p_i's must be approximated by multiples of $1/n$; here, a fine-structure study might be necessary.

2. Estimation of a single quantity. In this section we shall deal with the case where the interest of the experimenter is centered upon a particular linear combination of the parameters, say

$$\theta = a_1\beta_1 + a_2\beta_2. \tag{2.1}$$

Particularly we may have $\theta = \beta_1$ or $\theta = \beta_2$.

Consider all linear forms $t = \sum c_i\bar{y}_i$ yielding unbiased estimates of the quantity (2.1). For this purpose the c_i's have to make the equality

$$E(t) = \sum c_i(x_{i1}\beta_1 + x_{i2}\beta_2) = a_1\beta_1 + a_2\beta_2$$

identical in β_1, β_2, that is, they must satisfy the vector equation

$$\sum c_i \mathbf{x}_i = \mathbf{a}. \tag{2.2}$$

For $r > 2$ there are infinitely many sets $\{c_i\}$ fulfilling this condition. Among the corresponding estimates t there is, for every *fixed* set of p_i's, one with least variance; this statistic is, according to a theorem by Gauss, obtained by substituting in (2.1) the least-squares estimates $\hat{\beta}_1$, $\hat{\beta}_2$ of the parameters, that is, the values β_1, β_2 that minimize the weighted square sum

$$\sum p_i(\bar{y}_i - x_{i1}\beta_1 - x_{i2}\beta_2)^2. \tag{2.3}$$

The variance of this t is, of course, a function of $p_1, \cdots, p_r$. *We want to find those weights* p_i *which yield the smallest minimum variance.*

The solution of this problem can be found by a simple geometric argument. For this purpose we first notice that the smallest minimum variance, by definition, equals $\min_p \min_c D^2\{\sum c_i \bar{y}_i\}$, the c_i's and p_i's being subject to the conditions (2.2) and (1.2). Inverting the order of minimization and calculating, to begin with, the minimum with respect to the p_i's for fixed c_i's, we easily find

$$\min_p D^2 \{\textstyle\sum c_i \bar{y}_i\} = \min_p \sum \frac{c_i^2}{p_i} = k_c^2, \tag{2.4}$$

where $k_c = \sum |c_j|$. The minimizing p_i values are $p_{ci} = |c_i|/k_c$.

It remains to minimize (2.4) with respect to the c_i's, remembering the condition (2.2) which we rewrite in the form

$$\boldsymbol{a} = k_c \textstyle\sum p_{ci} \cdot \operatorname{sgn} c_i \cdot \boldsymbol{x}_i = k_c \boldsymbol{a}_c \text{ (say)}. \tag{2.5}$$

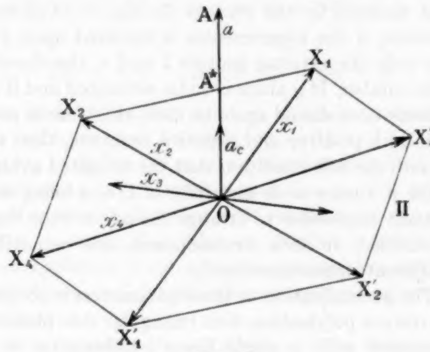


FIG. 1

The factor k_c being a positive scalar, the sum on the right-hand side represents a vector $\boldsymbol{a}_c$ with the same direction as $\boldsymbol{a}$. The weights p_{ci} being nonnegative with sum one, it is clear that the endpoint of this vector lies on or within the convex polygon Π spanned by the vectors $\pm\boldsymbol{x}_1, \cdots, \pm\boldsymbol{x}_r$ (Fig. 1). Since by (2.5), k_c is the length ratio of the vectors $\boldsymbol{a}$ and $\boldsymbol{a}_c$, it is obvious that (2.4) reaches its minimum when the endpoint of $\boldsymbol{a}_c$ coincides with the intersection A^* of the vector $\boldsymbol{a}$ (or its extension) with the polygon Π. If this point lies, for example, between the corners X_1 and X_2 of Π (see Fig. 1), it is seen that the coefficients p_{ci} in (2.5) —and hence the optimum weights—must be in the ratio $A^*X_2 : A^*X_1$ for $i = 1, 2$, and zero for $i = 3, \cdots, r$. The smallest minimum variance is given by $(OA/OA^*)^2$. In terms of our original problem we may state this result as follows:

For the estimation of a single quantity (2.1), *two and only two of the sources* (1.1) *are relevant; they have to be used in the proportion shown by the geometric argument above.* (There are obvious modifications of this statement in the cases where $\boldsymbol{a}$ passes through a corner of Π, or where three or more of the $\boldsymbol{x}_i$'s have their endpoints on the same line.)

If the vector of a certain source falls entirely inside the polygon spanned by the remaining vectors, this source is of no use for the estimation of any single quantity (2.1).

The fact that, with optimum allocation, only two of the potential observations are actually performed, has a somewhat surprising consequence. The square sum (2.3) reducing to two terms only, its absolute minimum, zero, is reached when β_1 and β_2 are chosen so as to make both terms vanish. The estimates $\hat{\beta}_1$, $\hat{\beta}_2$ are thus obtained simply by dropping the error terms in (1.3) and solving for the parameters.

Example. Consider the case where the potential observations are of form

$$y_i = \alpha + X_i\beta + \eta_i, \qquad X_1 < X_2 < \cdots < X_r,$$

that is, the case of linear regression in the elementary sense. Here the polygon Π is a parallellogram spanned by the vectors $(1, X_1)$, $(1, X_r)$, and their opposite vectors. If the interest of the experimenter is centered upon β alone, it is seen that he has to use only the extreme sources 1 and r; the observations on them have to be equal in number. If α alone is to be estimated and if all X_i's have the same sign, the extreme ones should again be used, this time in proportion $X_r : X_1$. If the X_i's include both positive and negative numbers, then the values of the p_i's are arbitrary with the sole condition that the weighted average $\sum p_i X_i = 0$. Since in practice the p_i's have to be multiples of $1/n$, n being the number of observations, it is usually impossible to arrange the p_i's so that the condition mentioned is exactly fulfilled. In such circumstances, one can still make a useful choice between different approximations[1]).

Generalization. The generalization to three parameters is obvious. The polygon Π is replaced by a convex polyhedron with triangular side-planes. In any estimation problem concerned with a single linear combination of the parameters there will in general be three relevant sources. For more than three parameters, the geometric rule must be replaced by an algebraic procedure.

3. Estimation of both parameters. For a set of actual observations, that is, for fixed $p_1, \cdots, p_r$, the least-squares technique yields minimum variance estimates of both parameters as well as of all linear combinations of them; nothing is gained in the accuracy of one estimate by giving up accuracy in another. In the present setup where the weights p_i are variable, some information is needed concerning the desired relative accuracy of different estimates. A reasonable approach seems to be to choose an appropriate positive definite quadratic form in the estimation errors $\hat{\beta}_1 - \beta_1$, $\hat{\beta}_2 - \beta_2$ and minimize its expectation by proper choice of the weights. By a linear transformation of the parameters (and a corresponding transformation of the coefficient vectors) the problem can always be reduced to the minimization of the particular form

$$q = E\{(\hat{\beta}_1 - \beta_1)^2 + (\hat{\beta}_2 - \beta_2)^2\} = D^2(\hat{\beta}_1) + D^2(\hat{\beta}_2) \tag{3.1}$$

with respect to $p_1, \cdots, p_r$.

[1] Cf. [1], p. 116. I am indebted to the referee for this and several other valuable remarks.

It is well known that the covariance matrix $\mathbf{\Lambda}$ of $\hat{\beta}_1$, $\hat{\beta}_2$ is the inverse of the information matrix

$$\mathbf{M} = \begin{bmatrix} \sum p_i x_{i1}^2 & \sum p_i x_{i1} x_{i2} \\ \sum p_i x_{i1} x_{i2} & \sum p_i x_{i2}^2 \end{bmatrix} = \sum p_i x_i' x_i . \tag{3.2}$$

Our object is to minimize the trace

$$q = \lambda_{11} + \lambda_{22} = \mu^{11} + \mu^{22}$$

of this inverse with respect to the p_i's. A peculiar feature of this problem is that the minimum point usually will lie on the boundary of the region (1.2), some of the p_i's being actually equal to zero.

Consider a point $P = (p_1, \cdots, p_r)$ in (1.2) in which q reaches its minimum. If i and j are two relevant sources, that is, if $p_i > 0$, $p_j > 0$, any differential variation

$$dp_i = -\delta, \qquad dp_j = \delta, \qquad dp_h = 0 \qquad (h \neq i, j)$$

of the coordinates leads to another point in (1.2). Accordingly, in order for P to be a minimum point, we must have $(\partial q/\partial p_j - \partial q/\partial p_i)\,\delta \geqq 0$ for *all* δ, that is, we must have $\partial q/\partial p_i = \partial q/\partial p_j$ for any two relevant sources i, j. If, on the other hand, i is a relevant and j an irrelevant source (i.e., $p_i > 0$, $p_j = 0$), then p_j can be varied only in the positive direction, and we must, by the same argument as above, have $(\partial q/\partial p_j - \partial q/\partial p_i)\delta \geqq 0$ for any *positive* δ; hence $\partial q/\partial p_j \geqq \partial q/\partial p_i$. In conclusion: to any solution of our minimization problem there exists a constant $-\kappa^2$ such that $\partial q/\partial p_i = -\kappa^2$ for all relevant sources, whereas $\partial q/\partial p_i \geqq -\kappa^2$ for irrelevant sources.

As far as the relevant sources are concerned, κ^2 is the ordinary Lagrange multiplier. Since q is a homogeneous function of order -1 of $p_1, \cdots, p_r$, and since, by the above result,

$$\sum_1^r p_i \frac{\partial q}{\partial p_i} = -\kappa^2 \sum_1^r p_i = -\kappa^2,$$

we conclude from Euler's identity that κ^2 equals the minimum value of q. This also establishes the sign of κ^2 as positive, as already anticipated in the notation.

We shall now compute the derivatives $\partial q/\partial p_i$. This is easily done by differentiating the matrix identity $\mathbf{M\Lambda} \equiv \mathbf{I}$ with respect to $p_i (i = 1, \cdots, r)$. Since by (3.2) $\partial\mathbf{M}/\partial p_i = x_i' x_i$, a short calculation gives

$$\frac{\partial \mathbf{\Lambda}}{\partial p_i} = -\mathbf{\Lambda} x_i' x_i \mathbf{\Lambda} = -(\mathbf{\Lambda} x_i')(\mathbf{\Lambda} x_i')'.$$

Hence

$$\frac{\partial q}{\partial p_i} = \frac{\partial \operatorname{sp} \mathbf{\Lambda}}{\partial p_i} = -\operatorname{sp}\{(\mathbf{\Lambda} x_i')(\mathbf{\Lambda} x_i')'\} = -\|\mathbf{\Lambda} x_i'\|^2, \tag{3.3}$$

where $\|\mathbf{\Lambda} x_i'\|$ denotes the length of the vector $\mathbf{\Lambda} x_i'$. We note that $\|\mathbf{\Lambda} x'\|^2$ is a

positive definite quadratic form in the components of $\mathbf{x}$; hence, the equation $\| \mathbf{A}\mathbf{x}' \|^2 = \text{Const.}$ represents an ellipse centered at the origin.

Combining the results of the three preceding paragraphs we have the following theorem.

THEOREM. *To any set* $\{p_i\}$ *that minimizes the function* (3.1) *there corresponds an ellipse* E, *centered at the origin, such that all points* $\mathbf{x}_i$ *representing relevant sources lie on* E *and none of the points representing irrelevant sources lie outside of* E.

Since three points determine a conic centered at the origin, we conclude that, in general, *there are at most three relevant sources.* Even in the case where four or more source-points happen to lie on the same ellipse, and the rest inside it, it may be shown by a continuity argument that three relevant sources are enough for the minimization of q.

Generalization. The preceding arguments apply to an arbitrary number s of parameters, the ellipse being replaced by an $(s - 1)$-dimensional ellipsoid or hyperellipsoid in R_s. Hence, there will be at most $\frac{1}{2}s(s + 1)$ relevant sources. However, already for $s = 3$ the computation of the optimum allocation becomes rather complicated.

4. Finding the weights. Simple examples show that the cases with two and with three relevant sources both actually occur. Assuming for the time being that we know how to pick these sources, we now want to find the weight distribution between them as well as the minimum value of q.

If there are two relevant sources corresponding, say, to $i = 1, 2$, we find the estimates $\hat{\beta}_1$, $\hat{\beta}_2$ simply by solving the equations $\bar{y}_i = x_{i1}\beta_1 + x_{i2}\beta_2$, $i = 1, 2$. Performing the solution, taking the variances, and introducing polar coordinates r, θ for the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, we find

$$q = D^2(\hat{\beta}_1) + D^2(\hat{\beta}_2) = \frac{r_2^2 p_1^{-1} + r_1^2 p_2^{-1}}{r_1^2 r_2^2 \sin^2(\theta_2 - \theta_1)}. \tag{4.1}$$

The minimum of this expression is attained for

$$p_1 = r_2/(r_1 + r_2), \qquad p_2 = r_1/(r_1 + r_2), \tag{4.2}$$

and the minimum value itself is

$$q_{\min} = \left\{\frac{r_1^{-1} + r_2^{-1}}{\sin(\theta_2 - \theta_1)}\right\}^2. \tag{4.3}$$

The case with three relevant sources $i = 1, 2, 3$ is somewhat more complicated. The derivatives (3.3) being less convenient for actual computation of the minimizing p_i values, we replace q by the homogeneous rational function L/M, where M is the determinant of the matrix (3.2) and L is the trace of M multiplied by $p_1 + p_2 + p_3$. On the set $p_1 + p_2 + p_3 = 1$ we obviously have $q = L/M$. Instead of minimizing L/M on the set mentioned, we may find the required ratio $p_1:p_2:p_3$ by minimizing L under the restriction $M = \text{Const.}$ Introducing a Lagrange multiplier λ and differentiating $L - \lambda M$ we find the equations

$$\sum_{j=1}^{3} (l_{ij} - \lambda m_{ij})p_j = 0 \qquad (i = 1, 2, 3), \tag{4.4}$$

where

$$l_{ij} = r_i^2 + r_j^2, \qquad m_{ij} = r_i^2 r_j^2 \sin^2(\theta_j - \theta_i). \tag{4.5}$$

If λ is any eigenvalue of the system (4.4) and P a corresponding eigenvector, a well known argument shows that λ is the value of L/M in P. Hence, *the minimum value of q on* (1.2) *is equal to the smallest eigenvalue of* (4.4) *for which all components of the eigenvector have same sign. The required optimum allocation is given by these components normalized to sum one.*

5. Selection of sources. There remains the question how to pick the relevant sources from a given set of potential observations.

From the theorem in Section 3 it is immediately seen that a source is certainly irrelevant if it is represented by a point inside the convex polygon spanned by all $\mathbf{x}$'s. A source i is, furthermore, irrelevant if for any two subscripts j, k ($\neq i$), the ellipse passing through $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$ and centered at the origin leaves some fourth $\mathbf{x}_h$ outside it. After discarding all such sources there might still be more than three left. It is in principle always possible to examine these "eligible sources" three by three, determine the corresponding minimum values of q according to Section 4, and pick the triplet with the smallest value. Most of the triplets will actually reduce to pairs, one of the three sources being irrelevant in combination with the others. The occurrence of this case is most easily detected by means of the following criterion, which we mention without proof:

A source $\mathbf{x}_3$ is irrelevant in combination with $\mathbf{x}_1$ and $\mathbf{x}_2$ if and only if $\mathbf{x}_3$ lies inside or on a certain ellipse, centered at the origin and passing through $\mathbf{x}_1$ and $\mathbf{x}_2$, with parametric equation

$$\mathbf{x} = \frac{\mathbf{x}_1 \sin(t - \theta_1) + \mathbf{x}_2 \sin(t - \theta_2)}{\sin(\theta_2 - \theta_1)} \qquad (0 \leq t \leq 2\pi). \tag{5.1}$$

In most practical situations, two sources picked by inspection will probably do without much loss of accuracy.

Example. Take three sources with polar coordinates (r, θ), $(r, -\theta)$, and (ρ, ϕ) respectively. We consider the two first as fixed, the third as variable. The equation of the ellipse (5.1) becomes in rectangular coordinates $\xi^2 \tan^2 \theta + \eta^2 \cot^2 \theta = r^2$. When $\mathbf{x}_3$ is inside this ellipse, the source 3 is irrelevant in combination with 1 and 2. There are, on the other hand, regions in which $\mathbf{x}_3$ "knocks out" one of the other sources. Writing (5.1) explicitly and interchanging the subscripts 1 (2) and 3 one finds, after some calculations, that source 1 (2) becomes irrelevant when $\mathbf{x}_3$ moves outside the curve $\rho \mid \sin 2\phi \mid = r \mid \sin 2\theta \mid$ in the first or third (second or fourth) quadrant. As a result we have a cross-shaped figure: in the center only sources 1, 2 are relevant, in the angle-fields only 2, 3 or 1, 3; along the axes all three sources are relevant.

6. Observations at different costs. The preceding theory can easily be adapted to the case where the potential observations are at different costs, say $\$v_1, \cdots, v_r$ per unit. Let $n_1, \cdots, n_r$ be the number of times that the different

observations are repeated. If the total costs have to equal a prescribed amount C, we have the restriction $\sum n_i v_i = C$ instead of $\sum n_i = n$. Dividing the regression equations for the averaged observations $\bar{y}_i$ by $\sqrt{v_i}(i = 1, \cdots, r)$ we get a new set of regression equations

$$\bar{y}_i^* = x_{i1}^* \beta_1 + x_{i2}^* \beta_2 + \frac{\eta_i^*}{\sqrt{p_i^*}} \qquad (i = 1, \cdots, r), \tag{6.1}$$

where

$$y_i^* = \bar{y}_i/\sqrt{v_i}, \qquad x_{ij}^* = x_{ij}/\sqrt{v_i}, \qquad p_i^* = v_i n_i/C, \tag{6.2}$$

where η_i^* is a random variable with mean zero and standard deviation $\sigma/\sqrt{C}$, and where the weights p_i^* are subject to the restrictions (1. 2). This is precisely the previous situation. One has only to enter the procedure with the modified sources x_i^* and to remember that the outcoming p_i^*'s give the optimum allocation of the costs, not of the observations themselves.

REFERENCES

[1] F. N. David and J. Neyman, "Extension of the Markoff theorem on least squares," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 105–116.

[2] H. Hotelling, "Some improvements in weighing and other experimental techniques," *Annals of Math. Stat.* Vol. 15 (1944), pp. 297–306.

ON THE DISTRIBUTION OF TWO RANDOM MATRICES USED IN CLASSIFICATION PROCEDURES[1]

By Rosedith Sitgreaves

Columbia University

Summary. Two classification statistics discussed in the literature can be written as functions of the elements of a $2 \cdot 2$ symmetric random matrix M. An analytic derivation is given of the distribution of M, and of a related matrix M^*, extending earlier work on distribution theory by Wald [1] and Anderson [2].

1. Introduction. A problem of classification discussed by Wald [1] and Anderson [2] may be described as follows. We have $N_1 + N_2 + 1$ independent p-dimensional chance vectors. We know that the first N_1 vectors are observations from a population π_1, the following N_2 are observations from a population π_2, and the last vector is an observation from a population π, where π is either π_1 or π_2. It is assumed that the probability distribution in both π_1 and π_2 is multivariate normal with the same covariance matrix Σ; the vector of expected values is $\mu^{(1)}$ in π_1 and $\mu^{(2)}$ in π_2. The values of $\mu^{(1)}$, $\mu^{(2)}$, and Σ are not known. Let X denote the $p \cdot (N_1 + N_2 + 1)$ matrix of observations. On the basis of X we want to classify the last observation, $X_{N_1+N_2+1}$ as coming from π_1 or π_2; that is, we want to make one of the two decisions, $\pi = \pi_1$ or $\pi = \pi_2$.

When the parameter values are known, the class of Bayes solutions is easily found, resulting in pairs of classification regions of the form

$$T^* \leqq k \quad \text{and} \quad T^* > k, \tag{1}$$

where

$$T^* = X'_{N_1+N_2+1}\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) - \tfrac{1}{2}(\mu^{(1)} + \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}). \tag{2}$$

Both Wald and Anderson propose, therefore, the use of classification statistics derived from (2) by substituting estimates for the unknown parameter values. Wald considers principally the statistic

$$U = X'_{N_1+N_2+1}S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}), \tag{3}$$

where

$$\bar{X}^{(1)} = (1/N_1)\sum_{t=1}^{N_1} X_t, \qquad \bar{X}^{(2)} = (1/N_2)\sum_{t=N_1+1}^{N_1+N_2} x_t,$$

and

$$S = (1/(N_1 + N_2 - 2)) \cdot \left[\sum_{t=1}^{N_1}(X_t - \bar{X}^{(1)})(X_t - \bar{X}^{(1)})' + \sum_{t=N_1+1}^{N_1+N_2}(X_t - \bar{X}^{(2)})(X_t - \bar{X}^{(2)})'\right].$$

[1] Sponsored by the Office of Naval Research.

Anderson proposes rather the statistic

$$(4) \qquad W = X'_{N_1+N_2+1}S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) - \tfrac{1}{2}(\bar{X}^{(1)} + \bar{X}^{(2)})'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}).$$

If we let $A = (N_1 + N_2 - 2)S$ and $[(N_1N_2)^{\frac{1}{2}}/(N_1 + N_2)^{\frac{1}{2}}](\bar{X}^{(1)} - \bar{X}^{(2)}) = Z$, we can write $U = [(N_1 + N_2)^{\frac{1}{2}}(N_1 + N_2 - 2)/(N_1N_2)^{\frac{1}{2}}]X'_{N_1+N_2+1}A^{-1}Z$. Under either alternative the vector variable Z has an expected value $[(N_1N_2)^{\frac{1}{2}}/(N_1 + N_2)^{\frac{1}{2}}](\mu^{(1)} - \mu^{(2)})$, and covariance matrix Σ. If $\pi = \pi_1$, the expected value of $X_{N_1+N_2+1}$ is $\mu^{(1)}$; if $\pi = \pi_2$, the expected value of $X_{N_1+N_2+1}$ is $\mu^{(2)}$. Thus, in either instance, the sampling distribution of U is contained as a special case of the sampling distribution of

$$(5) \qquad V = kY_1A^{-1}Y_2,$$

where k is a known scalar, Y_1 and Y_2 are p-dimensional normal variables with expected values ζ and ξ, say, respectively, and A is a $p \cdot p$ symmetric matrix with a Wishart distribution involving n degrees of freedom; the 3 sets of variables are independently distributed with the same covariance matrix Σ. Further, the statistic W can be written

$$\begin{aligned} W &= (X_{N_1+N_2+1} - (1/(N_1 + N_2))(N_1\bar{X}^{(1)} + N_2\bar{X}^{(2)}))'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \\ &\quad + ((1/(N_1 + N_2))(N_1\bar{X}^{(1)} + N_2\bar{X}^{(2)}) - \tfrac{1}{2}(\bar{X}^{(1)} - \bar{X}^{(2)}))'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \\ &= (X_{N_1+N_2+1} - (1/(N_1 + N_2))(N_1\bar{X}^{(1)} + N_2\bar{X}^{(2)}))'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \\ &\quad + [(N_1 - N_2)/(2N_1 + 2N_2)](\bar{X}^{(1)} - \bar{X}^{(2)})'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}). \end{aligned}$$

Or, if we let

$$[(N_1 + N_2)^{\frac{1}{2}}/(N_1 + N_2 + 1)^{\frac{1}{2}}] \cdot (X_{N_1+N_2+1} - (1/(N_1 + N_2))(N_1\bar{X}^{(1)} + N_2\bar{X}^{(2)})) = Z^*,$$

we can write

$$W = [(N_1 + N_2 + 1)^{\frac{1}{2}}(N_1 + N_2 - 2)/(N_1N_2)^{\frac{1}{2}}]Z^{*\prime}A^{-1}Z + [(N_1 - N_2)(N_1 + N_2 - 2)/(2N_1N_2)]Z'A^{-1}Z.$$

The vector variable Z^* is normally distributed, independently of Z, with covariance matrix Σ. Under the hypothesis $\pi = \pi_1$, the expected value of Z^* is $[N_2/(N_1 + N_2)^{\frac{1}{2}}(N_1 + N_2 + 1)^{\frac{1}{2}}](\mu^{(1)} - \mu^{(2)})$; under the hypothesis $\pi = \pi_2$, the expected value of Z^* is $[-N_1/(N_1 + N_2)^{\frac{1}{2}}(N_1 + N_2 + 1)^{\frac{1}{2}}](\mu^{(1)} - \mu^{(2)})$. The sampling distribution of W under either alternative is thus a special case of the sampling distribution of

$$(6) \qquad W^* = aY_1'A^{-1}Y_2 + bY_2'A^{-1}Y_2,$$

where a and b are known scalars, and Y_1, Y_2, and A are defined as before. In the case of W, the vectors ζ and ξ are proportional to $(\mu^{(1)} - \mu^{(2)})$.

Wald [1] investigated the general sampling distribution of V, and showed that the statistic can be expressed as a function of 3 variables. These variables, which

he called m_1, m_2, and m_3, and which become m_{11}, m_{22}, and m_{12} in our notation, are the elements of the symmetric matrix

$$M = Y'B^{-1}Y, \tag{7}$$

where $Y = (Y_1, Y_2)$ and $B = A + YY'$. The classification statistic V can be written

$$V = k \frac{m_{12}}{(1 - m_{11})(1 - m_{22}) - m_{12}^2}. \tag{8}$$

Wald showed geometrically that the distribution of M is a constant multiple of the product of 3 factors, the first a product of gamma- and beta-functions, the second an exponential term, and the third the expected value of a matrix of noncentral Wishart variables which was not evaluated. Anderson [2] has evaluated this product in the case when ζ and ξ are proportional.

In this paper, we give an analytic derivation of the distribution of M in the case when ζ and ξ are proportional, obtaining the constant of the distribution (which Wald and Anderson did not obtain). From the distribution of M, we obtain the distribution of the related matrix

$$M^* = Y'A^{-1}Y. \tag{9}$$

It can be easily shown that

$$M = M^*(I + M^*)^{-1}. \tag{10}$$

These distributions are useful because of interest in the exact sampling distributions of U and W. Further, as will be shown in a subsequent paper, an approach to the classification problem based on the principle of invariance results in a complete class of classification regions depending only on the elements of the matrix M^*, or equivalently of the matrix M, and on a single function of the parameters.

2. Distribution of M. We can write $\rho = k_1\delta$ and $\xi = k_2\delta$, where k_1 and k_2 are known scalars. The joint density function of Y and A is given by

$$p(Y,A) = \frac{|\Sigma|^{-\frac{1}{2}(n+2)} |A|^{\frac{1}{2}(n-p-1)}}{2^{\frac{1}{2}p(n+2)} \pi^{p+p(p-1)/4} \prod_{i=1}^{p} \Gamma(\frac{1}{2}(n+1-i))} \cdot \exp\{-\tfrac{1}{2}\lambda^2(k_1^2 + k_2^2) - \tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(A + YY') + \delta'\Sigma^{-1}(k_1Y_1 + k_2Y_2)\}, \tag{11}$$

where $\lambda^2 = \delta'\Sigma^{-1}\delta$. We make the transformation $B = A + YY'$. This is a one-to-one transformation with Jacobian 1. We have

$$p(Y,B) = \frac{|\Sigma|^{-\frac{1}{2}(n+2)} |B - YY'|^{\frac{1}{2}(n-p-1)}}{2^{\frac{1}{2}p(n+2)} \pi^{p+p(p-1)/4} \prod_{i=1}^{p} \Gamma(\frac{1}{2}(n+1-i))} \cdot \exp\{-\tfrac{1}{2}\lambda^2(k_1^2 + k_2^2) - \tfrac{1}{2} \operatorname{tr} \Sigma^{-1}B + \delta'\Sigma^{-1}(k_1Y_1 + k_2Y_2)\}. \tag{12}$$

There is a nonsingular matrix Ψ such that $\Psi\Sigma\Psi' = I$ and $\delta'\Psi' = (\lambda, 0, \cdots, 0)$ with $\lambda \geqq 0$. We make the transformation $Y^* = \Psi Y$ and $B^* = \Psi B\Psi'$. The Jacobian of the transformation is $|\Sigma|^{\frac{1}{2}(p+3)}$. Under the transformation

$$\begin{aligned}|B - YY'| &= |\Psi^{-1}B^*\Psi'^{-1} - \Psi^{-1}Y^*Y^{*\prime}\Psi'^{-1}|\\ &= |\Psi^{-1}(B^* - Y^*Y^{*\prime})\Psi'^{-1}| = |\Psi'\Psi|^{-1}|B^* - Y^*Y^{*\prime}|\\ &= |\Sigma|\cdot|B^* - Y^*Y^{*\prime}|,\end{aligned}$$

and

$$\delta'\Sigma^{-1}(k_1Y_1 + k_2Y_2) = \lambda(k_1y_{11}^* + k_2y_{12}^*).$$

Further,

$$\begin{aligned}M = Y'B^{-1}Y &= Y^{*\prime}\Psi'^{-1}(\Psi^{-1}B^*\Psi'^{-1})^{-1}\Psi^{-1}Y^*\\ &= Y^{*\prime}\Psi'^{-1}\Psi'B^{*-1}\Psi\Psi^{-1}Y^* = Y^{*\prime}B^{*-1}Y^*.\end{aligned}$$

The matrix B is positive definite with probability 1, and the matrix Ψ is nonsingular, so that the matrix B^* is positive definite with probability 1. We can write $B^* = TT'$, where T is a nonsingular triangular matrix whose elements are functions of the b_{ij}^*, chosen so that

$$t_{11} = b_{11}^{*\frac{1}{2}}, \qquad t_{ij} = 0 \qquad \text{for } j > i.$$

We use the matrix T to make the transformation $Y^* = TU$, where $U = (U_1, U_2)$ has the same dimensions as Y^*. The Jacobian of the transformation is $|T|^2 = |B^*|$. We have

$$\begin{aligned}|B^* - Y^*Y^{*\prime}| &= |TT' - TUU'T'| = |T(I - UU')T'|\\ &= |T|^2\cdot|I - UU'| = |B^*|\cdot|I - U'U|,\end{aligned}$$

since $|I - UU'| = |I - U'U|$. Also

$$M = Y^{*\prime}B^{*-1}Y^* = U'T'(TT')^{-1}TU = U'T'T'^{-1}T^{-1}TU = U'U.$$

The joint density function of U and B^* is given by

$$\begin{aligned}p(U, B^*) &= \frac{|B^*|^{\frac{1}{2}(n-p+1)}\,|I - U'U|^{\frac{1}{2}(n-p-1)}}{2^{\frac{1}{2}p(n+2)}\,\pi^{p+(p-1)/4}\prod_{i=1}^{p}\Gamma(\frac{1}{2}(n+1-i))}\\ &\quad\cdot\exp\{-\tfrac{1}{2}\lambda^2(k_1^2 + k_2^2) - \tfrac{1}{2}\operatorname{tr} B^* + \lambda b_{11}^{*\frac{1}{2}}(k_1u_{11} + k_2u_{12})\}\end{aligned} \tag{13}$$

The variables b_{ij}^* range over all values such that B^* is positive definite. The space of U is the set of points independent of the $b_{ij}^{*\prime}$'s for which

$$(1 - U_1'U_1) \geqq 0 \qquad (1 - U_2'U_2) \geqq 0 \qquad |U'U| \geqq 0$$

and

$$|I - U'U| \geqq 0.$$

It can be shown (e.g., see [3]) that

$$(14)\quad \underset{\substack{B^*_{(1)}\ \text{pos.def.} \\ -\infty \leq b^*_{1i}/b^{*\frac{1}{2}}_{11} \leq \infty}}{\int \cdots\cdots \int} | B^* |^{\frac{1}{2}(n-p+1)} e^{-\frac{1}{2}\operatorname{tr} B^*}\, db^*_{12} \cdots db^*_{pp}$$

$$= 2^{\frac{1}{2}(p-1)(n+2)} \pi^{p(p-1)/4} b_{11}^{*\frac{1}{2}n} e^{-\frac{1}{2}b^*_{11}} \prod_{i=1}^{p-1} \Gamma(\tfrac{1}{2}(n+2-i)), \quad i, j = 2, \cdots, p,$$

where $B^*_{(1)} = (b^*_{ij} - b^*_{1i} b^*_{1j}/b^*_{11})$. Hence

$$(15)\quad p(U, b^*_{11}) = \frac{\Gamma(\frac{1}{2}(n+1)) \mid I - U'U \mid^{\frac{1}{2}(n-p-1)} b_{11}^{*\frac{1}{2}n}}{\Gamma(\frac{1}{2}(n-p+2))\Gamma(\frac{1}{2}(n-p+1)) 2^{\frac{1}{2}(n+2)} \pi^p}$$

$$\cdot \exp\{-\tfrac{1}{2}b^*_{11} - \tfrac{1}{2}\lambda^2(k_1^2 + k_2^2) + \lambda b_{11}^{*\frac{1}{2}}(k_1 u_{11} + k_2 u_{12})\}.$$

Expanding $\exp(\lambda b_{11}^{*\frac{1}{2}}(k_1 u_{11} + k_2 u_{12}))$ in a power series and integrating with respect to b^*_{11} we obtain

$$(16)\quad p(U) = \frac{\Gamma(\frac{1}{2}(n+1)) \mid I - U'U \mid^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\lambda^2(k_1^2+k_2^2)}}{\Gamma(\frac{1}{2}(n-p+2))\Gamma(\frac{1}{2}(n-p+1))\pi^p}$$

$$\cdot \sum_{j=0}^{\infty} \frac{\Gamma(\frac{1}{2}(n+2+j)) 2^{\frac{1}{2}j} \lambda^j (k_1 u_{11} + k_2 u_{12})^j}{j!}.$$

We can construct an orthogonal matrix G as follows. Let

$$g_{1j} = u_{j1} \Big/ \Big(\sum_{l=1}^{p} u_{l1}^2\Big)^{\frac{1}{2}} \qquad j = 1, 2, \cdots, p,$$

$$g_{21} = -\Big(\sum_{l=2}^{p} u_{l1}^2\Big)^{\frac{1}{2}} \Big/ \Big(\sum_{l=1}^{p} u_{l1}^2\Big)^{\frac{1}{2}} \qquad g_{2j} = u_{11} u_{j1} \Big/ \Big(\sum_{l=1}^{p} u_{l1}^2\Big)^{\frac{1}{2}} \Big(\sum_{l=2}^{p} u_{l1}^2\Big)^{\frac{1}{2}},$$

$$j = 2, \cdots, p.$$

For $k = 3, \cdots, p-1$

$$g_{kj} = 0, \quad j = 1, \cdots, k-2; \qquad g_{k,k-1} = -\Big(\sum_{l=k}^{p} u_{l1}^2\Big)^{\frac{1}{2}} \Big/ \Big(\sum_{l=k-1}^{p} u_{l1}^2\Big)^{\frac{1}{2}}$$

$$g_{kj} = u_{k-1\,1} u_{j1} \Big/ \Big(\sum_{l=k-1}^{p} u_{l1}^2\Big)^{\frac{1}{2}} \Big(\sum_{l=k}^{p} u_{l1}^2\Big)^{\frac{1}{2}}, \quad j = k, \cdots, p;$$

$$g_{pj} = 0, \quad j = 1, \cdots, p-2; \qquad g_{p,p-1} = -u_{p1}/(u_{p-1,1}^2 + u_{p1}^2)^{\frac{1}{2}};$$

$$g_{pp} = u_{p-1,1}/(u_{p-1,1}^2 + u_{p1}^2)^{\frac{1}{2}}.$$

We make the transformation $V = GU_2$. Under the transformation

$$\begin{aligned} | I - U'U | &= | (1 - U_1'U_1)(1 - U_2'U_2) - U_1'U_2U_2'U_1 | \\ &= | (1 - U_1'U_1)(1 - V'V) - v_1^2 U_1'U_1 | \end{aligned}$$

and

$$u_{12} = v_1 u_{11} \Big/ \Big(\sum_{l=1}^{p} u_{l1}^2\Big)^{\frac{1}{2}} - v_2 \Big(\sum_{l=2}^{p} u_{l1}^2\Big)^{\frac{1}{2}} \Big/ \Big(\sum_{l=1}^{p} u_{l1}^2\Big)^{\frac{1}{2}}.$$

Now we make the following transformation. Let

$$
\begin{aligned}
u_{11} &= m_{11}^{\frac{1}{2}} \cos \theta_1 , \\
u_{21} &= m_{11}^{\frac{1}{2}} \sin \theta_1 \cos \theta_1 , \\
&\cdot \\
&\cdot \\
&\cdot \\
u_{p-1,1} &= m_{11}^{\frac{1}{2}} \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{p-2} \cos \theta_{p-1} , \\
u_{p1} &= m_{11}^{\frac{1}{2}} \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{p-2} \sin \theta_{p-1} .
\end{aligned}
$$

The Jacobian of the transformation is

$$\tfrac{1}{2} m_{11}^{\frac{1}{2}(p-2)} \sin \theta_1^{p-2} \sin \theta_2^{p-3} \cdots \sin \theta_{p-2}$$

with $0 \leqq \theta_i \leqq \pi$ for $i = 1, 2, \cdots, p - 2$ and $0 \leqq \theta_{p-1} \leqq 2\pi$.

Under the transformation, $U'_1U_1 = m_{11}$ and

$$
\begin{aligned}
&p(m_{11}, V, \theta) \\
&= \frac{\Gamma(\frac{1}{2}(n+1)) e^{-\frac{1}{2}\lambda^2(k_1^2+k_2^2)} m_{11}^{\frac{1}{2}(p-2)} \left((1 - m_{11})(1 - \sum_{i=1}^{p} v_i^2) - m_{11} v_1^2 \right)^{\frac{1}{2}(n-p-1)}}{2\pi^p \Gamma(\frac{1}{2}(n - p + 2)) \Gamma(\frac{1}{2}(n - p + 1))} \\
&\qquad \cdot \sin \theta_1^{p-2} \cdots \sin \theta_{p-2} \\
&\qquad \cdot \sum_{j=0}^{\infty} \frac{\Gamma(\frac{1}{2}(n + 2 + j)) 2^{\frac{1}{2}j} \lambda^j (k_1 m_{11}^{\frac{1}{2}} \cos \theta_1 + k_2 v_1 \cos \theta_1 - k_2 v_2 \sin \theta_1)^j}{j!} .
\end{aligned} \tag{17}
$$

Since

$$\int_0^{\frac{1}{2}\pi} \sin^m \theta \cos^n \theta \, d\theta = \frac{1}{2} \frac{\Gamma(\frac{1}{2}(m+1)) \Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}(m+n) + 1)} ,$$

it follows that

$$\int_0^{\pi} \sin^{p-i-1} \theta_i \, d\theta_i = 2 \int_0^{\frac{1}{2}\pi} \sin^{p-i-1} \theta_i \, d\theta_i = \frac{\Gamma(\frac{1}{2}) \Gamma(\frac{1}{2}(p - i))}{\Gamma(\frac{1}{2}(p - i + 1))}, \quad i = 2, \cdots, p - 2,$$

and

$$(\Gamma(\tfrac{1}{2}))^{p-3} \prod_{i=2}^{p-2} \frac{\Gamma(\frac{1}{2}(p - i))}{\Gamma(\frac{1}{2}(p - i + 1))} = \frac{\pi^{\frac{1}{2}(p-3)}}{\Gamma(\frac{1}{2}(p-1))} .$$

Further, $\int_0^{2\pi} d\theta_{p-1} = 2\pi$ so that we have

$$
\begin{aligned}
&p(m_{11}, V, \theta_1) \\
&= \frac{\Gamma(\frac{1}{2}(n+1)) e^{-\frac{1}{2}\lambda^2(k_1^2+k_2^2)} m_{11}^{\frac{1}{2}(p-2)} \left((1 - m_{11})(1 - \sum_{i=1}^{p} v_i^2) - m_{11} v_1^2 \right)^{\frac{1}{2}(n-p-1)}}{\Gamma(\frac{1}{2}(n - p + 2)) \Gamma(\frac{1}{2}(n - p + 1)) \Gamma(\frac{1}{2}(p - 1)) \, \pi^{\frac{1}{2}(p+1)}} \\
&\qquad \cdot \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \frac{\Gamma(\frac{1}{2}(n + 2 + j + l))}{j! \, l!} \\
&\qquad \cdot (2^{\frac{1}{2}}\lambda)^{j+l} (k_1 m_{11}^{\frac{1}{2}} + k_2 v_1)^j (-k_2 v_2)^l \sin \theta_1^{p-2+l} \cos \theta_1^j .
\end{aligned} \tag{18}
$$

Since $\int_0^{\pi} \sin^m \theta \cos^n \theta \, d\theta = 0$ for n odd, we obtain on integrating with respect to θ_1

$$
\begin{aligned}
&p(m_{11}, V) \\
&= \frac{\Gamma(\frac{1}{2}(n+1))e^{-\frac{1}{2}\lambda^2(k_1^2+k_2^2)} m_{11}^{\frac{1}{2}(p-2)} \left((1-m_{11})(1-\sum_{i=1}^{p} v_i^2) - m_{11}v_1^2\right)^{\frac{1}{2}(n-p-1)}}{\Gamma(\frac{1}{2}(n-p+2))\Gamma(\frac{1}{2}(n-p+1))\Gamma(\frac{1}{2}(p-1))\pi^{\frac{1}{2}(p+1)}} \\
&\quad \cdot \sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\left\{\frac{\Gamma(\frac{1}{2}(n+2+l)+j)\Gamma(\frac{1}{2}(p-1+l))\Gamma(j+\frac{1}{2})2^{j+\frac{1}{2}l}\lambda^{2j+l}}{\Gamma(\frac{1}{2}(p+l)+j)(2j)!\,l!}\right. \\
&\qquad\qquad \left.\cdot (k_1 m_{11}^{\frac{1}{2}} + k_2 v_1)^{2j}(-k_2 v_2)^l\right\}.
\end{aligned} \tag{19}
$$

We partition the vector V into two parts, the first part consisting of the single element v_1, and the second part of the $(p-1)$-dimensional vector V^*. In a manner similar to that in which U_1 was transformed, we transform the vector V^* to a variable $m_{22.1} = V^{*\prime}V^*$, and to $(p-2)$ angles. After integrating with respect to the angles, and simplifying the resulting expression, we obtain

$$
\begin{aligned}
p(m_{11}, m_{22.1}, v_1) &= \frac{\Gamma(\frac{1}{2}(n+1))e^{-\frac{1}{2}\lambda^2(k_1^2+k_2^2)} m_{11}^{\frac{1}{2}(p-2)} m_{22.1}^{\frac{1}{2}(p-3)}}{\Gamma(\frac{1}{2}(n-p+2))\Gamma(\frac{1}{2}(n-p+1))\Gamma(\frac{1}{2}(p-1))\Gamma(\frac{1}{2})} \\
&\quad \cdot ((1-m_{11})(1-m_{22.1}) - v_1^2))^{\frac{1}{2}(n-p-1)} \\
&\quad \cdot \sum_{j=0}^{\infty} \frac{\Gamma(\frac{1}{2}(n+2)+j)}{\Gamma(\frac{1}{2}p+j)\,j!}(\tfrac{1}{2}\lambda^2)^j (k_1^2 m_{11} + 2k_1k_2 m_{11}^{\frac{1}{2}} v_1 + k_2^2(v_1^2 + m_{22.1}))^j.
\end{aligned} \tag{20}
$$

Finally, we make the transformation

$$v_1 = m_{12}/m_{11}^{\frac{1}{2}} \qquad m_{22.1} = m_{22} - m_{12}^2/m_{11}$$

and we have

$$
\begin{aligned}
p(M) &= \frac{\Gamma(\frac{1}{2}(n+1))e^{-\frac{1}{2}\lambda^2(k_1^2+k_2^2)} |M|^{\frac{1}{2}(p-3)} |I-M|^{\frac{1}{2}(n-p-1)}}{\Gamma(\frac{1}{2}(n-p+2))\Gamma(\frac{1}{2}(n-p+1))\Gamma(\frac{1}{2}(p-1))\Gamma(\frac{1}{2})} \\
&\quad \cdot \sum_{j=0}^{\infty} \frac{\Gamma(\frac{1}{2}(n+2)+j)}{\Gamma(\frac{1}{2}p+j)j!}(\tfrac{1}{2}\lambda^2)^j (k_1^2 m_{11} + 2k_1k_2 m_{12} + k_2^2 m_{22})^j,
\end{aligned} \tag{21}
$$

with $0 \leqq m_{11} \leqq 1$, $0 \leqq m_{22} \leqq 1$, $|M| \geqq 0$, $|I - M| \geqq 0$.

3. Distribution of M^*. Making the transformation defined by $M = M^*(I + M^*)^{-1}$, we obtain

$$
\begin{aligned}
p(M^*) &= \frac{\Gamma(\frac{1}{2}(n+1))e^{-\frac{1}{2}\lambda^2(k_1^2+k_2^2)} |M^*|^{\frac{1}{2}(p-3)}}{\Gamma(\frac{1}{2}(n-p+2))\Gamma(\frac{1}{2}(n-p+1))\Gamma(\frac{1}{2}(p-1))\Gamma(\frac{1}{2})} \\
&\quad \cdot \sum_{j=0}^{\infty}\left\{\frac{\Gamma(\frac{1}{2}(n+2)+j)}{\Gamma(\frac{1}{2}p+j)j!}(\tfrac{1}{2}\lambda^2)^j\right. \\
&\qquad \left.\cdot \frac{(k_1^2 m_{11}^* + 2k_1k_2 m_{12}^* + k_2^2 m_{22}^* + (k_1^2+k_2^2)(m_{11}^* m_{22}^* - m_{12}^{*2}))^j}{|I+M^*|^{\frac{1}{2}(n+2)+j}}\right\}.
\end{aligned} \tag{22}
$$

Acknowledgment. Grateful acknowledgment is made to Prof. T. W. Anderson, under whose direction this paper was prepared.

REFERENCES

[1] A. Wald, "On a statistical problem arising in the classification of an individual into one of two groups," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 145–162.

[2] T. W. Anderson, "Classification by multivariate analysis," *Psychometrika*, Vol. 16 (1951), pp. 31–50.

[3] T. W. Anderson and M. A. Girshick, "Some extensions of the Wishart distribution," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 345–357.

THE DISTRIBUTION OF THE NUMBER OF ISOLATES IN A SOCIAL GROUP[1]

By Leo Katz

University of North Carolina and Michigan State College

1. Summary. The exact chance distribution of the number of isolates in a social group is found in this paper, using methods due to Fréchet. The binomial distribution fitted to the first two moments of the exact distribution is shown to give reasonably good approximation and a slightly coarser binomial approximation is indicated.

2. Introduction. Consider a group consisting of N individuals. Each designates d of the others with whom he would prefer to be associated in some specified activity, that is, each chooses d from $N - 1$ possible associates. In the context of the group and the specified activity, an individual is said to be an *isolate* if he is chosen by none of his fellow group members. It is immediately obvious that the number of isolates depends upon the size of the group, the number of choices permitted and the extent to which the group, as a social organism, provides acceptance for joint activities for the individuals who compose the group. Thus, when N and d are fixed, the number of isolates becomes an important characteristic of the group structure. When it is important to state whether the number of isolates is unusually large or small, it is necessary that the chance distribution of this number be known.

The history of attacks on the distribution problem is brief. Lazarsfeld, in a contribution to a paper by Moreno and Jennings [8], gave the expected (mean) number of isolates as

$$N[(N - d - 1)/(N - 1)]^{N-1},$$

but made no attempt to obtain the distribution. Bronfenbrenner [1] gave (without proof) an incorrect version of the distribution function. He gave the expression, which he claimed was "developed deductively and checked by empirical methods,"

$$P(i) = \Pr\{i \text{ or fewer isolates}\} = 1 - \frac{(N - i - 2)^{(d)}}{(N - 1)^{(d)}}, \tag{1}$$

where $a^{(b)} = a(a - 1)(a - 2)\cdots(a - b + 1)$. This form gives completely nonsensical results in application. Edwards [2] conjectured that the Bronfenbrenner formula gives the probability of a given person's including in his list of d at least

[1] Work done under the sponsorship of the Office of Naval Research at Chapel Hill, North Carolina, and presented at the Chicago meeting of the Institute of Mathematical Statistics, December 27, 1950.

one of $i + 1$ specified names. Edwards then gave correctly the probability of the maximum possible number of isolates,

$$(2)\quad P[N-1-d] = \Pr\{N-1-d \text{ isolates}\} = \binom{N}{N-1-d}\frac{(d+1)^{(N-1-d)}}{\binom{N-1}{d}^{N}},$$

where $\binom{a}{b}$, $b \leq a$, is the binomial coefficient $a!/[b!(a-b)!]$. Note that there cannot be $N - d$ isolates, since d persons can be chosen only for a maximum total of $(N - 1)d$ times, less than the Nd choices actually made.

In the last paper cited above, Edwards went on to set up the probability of $N - 2 - d$ isolates by eliminating irrelevant cases from those in which the isolates name d from a list of $d + 2$ while the nonisolates choose d from a list of $d + 1$ names, and indicated that the process might be continued to obtain the probabilities of $N - 3 - d$ isolates, etc. The form of these results, it is stated, would indicate a complicated algebraic expression for the required probability distribution and the question is then raised whether the existing technique of experimentation should not be modified to meet the practical requirement of simple mathematical treatment.

In this paper, we shall first obtain the exact distribution of the number of isolates on the assumption of random choice and second, we shall obtain an approximation which *does* satisfy the requirement of simple mathematical treatment. An example will be given to indicate the accuracy of the approximation for a typical application.

3. Exact distribution of the number of isolates. It should first be remarked that any division of the group into those who are isolates and those who may not be produces two distinct patterns of choices. Each isolate selects d from among all those in the second group, but each member of the second group must select d from among those members of the second group not including himself. Let

$$p_{i_1,i_2,\cdots,i_k} = \Pr\{\text{individuals } i_1, i_2, \cdots, i_k \text{ are isolates}\}.$$

As an immediate consequence of the remark made above and the symmetry of the situation,

$$(3)\qquad p_{i_1,i_2,\cdots,i_k} = \left[\frac{\binom{N-k}{d}}{\binom{N-1}{d}}\right]^{k}\cdot\left[\frac{\binom{N-k-1}{d}}{\binom{N-1}{d}}\right]^{N-k},$$

for every $(i_1, i_2, \cdots, i_k)$. Setting

$$(4)\quad S_k = \binom{N}{k}p_{i_1,i_2,\cdots,i_k} = \binom{N}{k}\binom{N-k}{d}^{k}\binom{N-k-1}{d}^{N-k}\binom{N-1}{d}^{-N}$$

the principle of inclusion and exclusion ([3], ch. 4) gives immediately

$$(5)\quad P_{[k]} = \Pr\{\text{exactly } k \text{ isolates in the group}\} = \sum_{j=k}^{N-1-d}(-1)^{k+j}\binom{j}{k}S_j.$$

Equation (5) gives the exact probability of k isolates, in a group of N where each individual makes d choices, as a linear combination of the S_k.

The values of S_k may be computed directly from (4) or recursively, noting that $S_0 = 1$ and

$$\frac{S_{k+1}}{S_k} = \frac{N-k-d}{k+1}\left[\frac{N-k-d}{N-k}\right]^{k-1}\left[\frac{N-k-1-d}{N-k-1}\right]^{N-k-1}. \tag{6}$$

The form of the last term in (6) suggests interesting asymptotic behavior. We are, however, less interested in asymptotic characteristics of the distribution than in its properties for moderate values of N. We may take the asymptotic behavior to give an indication of what may be a reasonable approximation, but the quality of the approximation must be judged by results for typical cases; here, N is usually between 10 and 100. We shall later consider one such typical example in which $N = 26$, $d = 3$.

If we do not require the values of the individual $P_{[i]}$ but are only interested in the moments of the distribution of isolates, it turns out that the S_k quantities are of central importance. Fréchet [4] has shown that

$$\alpha_{(k)} = k!\, S_k, \tag{7}$$

where $\alpha_{(k)}$ is the kth factorial moment of the distribution, given by $\alpha_{(k)} = \sum_{i=1}^{N-1-d} i^{(k)} P_{[i]}$. We shall have occasion to use these factorial moments in the following section.

4. Approximate distribution of number of isolates. Since we know the exact distribution, an approximate distribution is useful only if it is more easily computed. It is easily shown (see Feller [3]) that, for d fixed, the limiting distribution is Poisson with $\Pr(k) = e^{-\lambda}\lambda^k/k!$, where $\lambda = N(1 - d/(N-1))^{N-1}$. However, for moderate values of N, the approximation is not good; an example is given later.

Following the procedure of Kaplansky [7] produces a modified Poisson approximation which is quite good. The drawback to this procedure is that computations are almost as difficult as for the exact distribution. Therefore, we seek another approximation to satisfy the dual requirements of accuracy and simplicity.

From (4) and (7), the mean and the variance of the number of isolates are respectively,

$$\text{mean} = \alpha_{(1)} = N\left(1 - \frac{d}{N-1}\right)^{N-1}, \tag{8}$$

$$\begin{aligned}\text{variance} &= \alpha_{(2)} + \alpha_{(1)} - \alpha_{(1)}^2 \\ &= N\left(1 - \frac{d}{N-1}\right)^{N-1}\left[1 + (N-1-d)\left(1 - \frac{d}{N-2}\right)^{N-2} - N\left(1 - \frac{d}{N-1}\right)^{N-1}\right].\end{aligned} \tag{9}$$

From (9), we see var (k) = mean (k) $[1 - (d+1)(1 - d/(N-2))^{N-2} + O(N^{-2})]$ $\approx$ mean (k) $[1 - (d+1)e^{-d}]$. Since the variance is less than the mean, the binomial distribution, $b(x; n, p)$, is strongly suggested (choice being restricted to simple distributions). We shall not insist that n be an integer; thus, we have essentially a beta distribution. For this distribution, $\alpha_{(r)} = n^{(r)}p^r$ and, fitting the first two moments, we have

$$np = \alpha_{(1)} = N\left(1 - \frac{d}{N-1}\right)^{N-1}, \tag{10}$$

$$\frac{1}{n} = 1 - \frac{\alpha_{(2)}}{\alpha_{(1)}^2} = 1 - \left(1 - \frac{1}{N}\right)\left[1 - \frac{d}{(N-2)(N-1-d)}\right]^{N-2}. \tag{11}$$

Also, since $\alpha_{(r+1)}/\alpha_{(r)} = (n - r)p$, we form the functions,

$$D_r = \frac{\alpha_{(r+1)}}{\alpha_{(r)}} - r\frac{\alpha_{(2)}}{\alpha_{(1)}} + (r-1)\alpha_{(1)}, \qquad r = 2, 3, 4, \cdots, \tag{12}$$

which vanish identically for the binomial distribution. These functions are equivalent to the "total criteria" proposed by Guldberg [6] and Frisch [5] for judging whether an observed series may be approximated by a binomial frequency function. In their work, the approximation is considered to be good when the criterion functions of the moments of the observed series are close to zero. We shall extend the notion to cover the case of approximation of a more complicated probability law by the binomial law.

Setting $r = 2$ and $r = 3$ in (12) gives two functions which are exactly equivalent to the two criteria given by Guldberg (allowing for an omitted term in his second result). Also, the complete set (12) is equivalent to Frisch's total criteria for $g = 1$, $h = 1, 2, 3, \cdots$ in his notation. Since his criteria for all other values of g may be expressed in terms of those for $g = 1$, (12) is equivalent to the complete set of conditions given by Frisch.

Substituting from equation (7) into (12), we have

$$D_r = (r+1)\frac{S_{r+1}}{S_r} - 2r\frac{S_2}{S_1} + (r-1)S_1,$$

or, using (4) and (6),

$$\begin{aligned} D_r = {} & (N - r - d)\left(\frac{N-r-d}{N-1}\right)^{r-1}\left(\frac{N-r-1-d}{N-r-1}\right)^{N-r-1} \\ & - r(N-1-d)\left(\frac{N-2-d}{N-2}\right)^{N-2} + N(r-1)\left(\frac{N-1-d}{N-1}\right)^{N-1}. \end{aligned} \tag{13}$$

For large N, each power of a fraction in (13) of the form $((a-d)/a)^a$ is approximately equal to e^{-d} and $D_r = 0$, approximately. In the limit, every $D_r = 0$; the asymptotic form of the distribution in this sense is, therefore, binomial. Further, the approximation should remain good even for moderate values of N (particularly when r is small) since the errors made by the exponential approximation are not only small but tend to compensate for each other.

We may, then, use a binomial probability law approximation with p and n given by (10) and (11). (If $1/n$ in (11) is evaluated to terms of $O(N^{-2})$, we find $1/n = (d + 1)/(N - 1 - d)$ or $n = N/(d + 1) - 1$, approximately. This seems consistently to understate the value of n from (11); accordingly, it is suggested that n be approximated by

$$n = \frac{N}{d+1} - \frac{1}{2}. \tag{11a}$$

In the next section, we shall compare this approximation with the exact distribution for a typical pair of values of N and d. We also give, for comparison, the Poisson approximation.

TABLE 1

Comparison of the exact and approximate distributions of the number of isolates for $N = 26, d = 3$

i	S_i	$P_{[i]}$(exact)	p_i(approx.)	$p_i - P_{[i]}$	$p'_i = \frac{e^{-\lambda}\lambda^i}{i!}$	$p'_i - P_{[i]}$
0	1.000 0000	.309 794	.311 098	+.0013	.344 989	+.0352
1	1.064 2429	.402 574	.399 727	−.0028	.367 152	−.0354
2	.474 9281	.214 316	.215 365$^+$	+.0010	.195 370	−.0189
3	.116 8650$^+$	.061 532	.062 473	+.0009	.069 306	+.0078
4	.017 5606	.010 564	.010 354	−.0002	.018 440	+.0079
5	.001 6882	.001 138	.000 943	−.0002	.003 925$^-$	+.0028
6	.000 10596	.000 079	.000 039	−.00004	.000 696	+.00062
7	.000 0043 61	.000 003	.000 0002	−.000003	.000 106	+.000103
8	.000 0001 17				.000 014	
9	.000 0000 02				.000 002	

5. An example. Moreno and Jennings [8] considered in some detail the case $N = 26$, $d = 3$. Since, also, a number of later writers have treated the same case as a reasonably typical one, we will test the accuracy of the approximation in this situation. The computation of the exact probability distribution seems to be best performed in two stages. In the first, the logarithms of the ratios S_{j+1}/S_j of equation (6) are obtained using 7-place tables, and the S_i themselves obtained from the partial sums of the logarithms. These values appear in the second column of Table 1. In the second stage of the computation, the exact probabilities are found by setting the S_i into (5). The exact probabilities are given to six decimals in the third column of the table.

In the computation of the approximate probabilities, we take advantage of the already computed values of S_1 and S_2 and equation (7) to obtain directly the factorial moments of (8) and (9). From (10) and (11), we have $p = .1717247$ and $n = 6.197378$. We then compute the binomial probabilities, $p_i = b(i; n, p)$,

$i = 0, 1, 2, \cdots, ([n] + 1)$, where $[n]$ is the largest integer in n, in this case, 6, using $p_0 = (1 - p)^n$ and $p_{i+1}/p_i = (n - i)p/(i + 1)(1 - p)$ as suggested by Guldberg [6] and others. The approximate probabilities, p_i, appear in the fourth column of the table to six decimals. It will be seen that the fit to three decimals is almost exact and certainly good enough for tests of significance. The discrepancies, $p_i - P_{[i]}$, are given in the fifth column. The Poisson probabilities and errors appear in the sixth and seventh columns.

The discrepancies for the "binomial" approximation are not particularly systematic except in the upper tail of the distribution, where the binomial gives zero probability for all numbers of isolates above seven. Although numbers through 22 are possible, they are so unlikely to occur by chance that this possibility may be practically disregarded. For example, the exact probability of eight isolates by chance is about one in ten million. The Poisson distribution appears to be "flatter" than the exact, understating probabilities for the central values and overstating for both tails.

As a further check on the accuracy of the approximation, the values of $\gamma_1 = \mu_3/\mu_2^{3/2}$ and $\gamma_2 = \mu_4/\mu_2^2$ were computed for the exact distribution and for the "binomial" approximation. These computations give $\gamma_1 = .7193$ for the exact, .6993 for the approximate distribution; $\gamma_2 = 3.2620$ and 3.1663, respectively.

REFERENCES

[1] U. Bronfenbrenner, "The measurement of sociometric status, structure and development," *Sociometry*, Vol. 6 (1943), pp. 363–397. Reprinted as *Sociometry Monograph No. 6*, Beacon House, New York, 1945.

[2] D. S. Edwards, "The constant frame of reference problem in sociometry," *Sociometry*, Vol. 11 (1948), pp. 372–379.

[3] W. Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, 1950.

[4] M. Fréchet, *Les Probabilités Associées à un système d'Événements Compatibles et Dépendants, Actualités Scientifiques et Industrielles, Nos. 859 and 942*, Hermann et Cie, Paris, 1940 and 1943.

[5] R. Frisch, "On the use of difference equations in the study of frequency distributions," *Metron*, Vol. 10 (1932), pp. 35–59.

[6] A. Guldberg, "On discontinuous frequency functions and statistical series," *Skandinavisk Aktuarietidskrift*, Vol. 14 (1931), pp. 161–187.

[7] I. Kaplansky, "The asymptotic distribution of runs of consecutive elements," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 200–203.

[8] J. L. Moreno and H. H. Jennings, "Statistics of social configurations," *Sociometry*, Vol. 1 (1938), pp. 342–374. Reprinted as *Sociometry Monograph No. 3*, Beacon House, New York, 1945.

NOTES

JUSTIFICATION AND EXTENSION OF DOOB'S HEURISTIC APPROACH TO THE KOLMOGOROV-SMIRNOV THEOREMS[1]

By Monroe D. Donsker

University of Minnesota

1. Introduction and summary. Doob [1] has given heuristically an appealing methodology for deriving asymptotic theorems on the difference between the empirical distribution function calculated from a sample and the actual distribution function of the population being sampled. In particular he has applied these methods to deriving the well known theorems of Kolmogorov [2] and Smirnov [3]. In this paper we give a justification of Doob's approach to these theorems and show that the method can be extended to a wide class of such asymptotic theorems.

2. The justification for Kolmogorov's theorem. Let $x_1, x_2, \cdots$, be mutually independent, identically distributed random variables with distribution function $F(\lambda)$, and let $\nu_n(\lambda)$ be the number of x_i's among $x_1, x_2, \cdots, x_n$ which are $\leqq \lambda$. In studying the difference between the empirical distribution function, $\nu_n(\lambda)/n$, and $F(\lambda)$, Kolmogorov showed that if $F(\lambda)$ is continuous, the distribution of

$$\operatorname*{l.u.b.}_{-\infty<\lambda<+\infty}\left(\frac{\nu_n(\lambda)}{n} - F(\lambda)\right) \tag{2.1}$$

is independent of $F(\lambda)$. For convenience, therefore, we will assume that the variables are uniformly distributed on (0, 1), that is, $F(\lambda) = \lambda$ for $0 \leqq \lambda \leqq 1$. Let[2]

$$D_n^+ = \operatorname*{l.u.b.}_{0\leqq\lambda\leqq 1}\left(\frac{\nu_n(\lambda)}{n} - \lambda\right). \tag{2.2}$$

One of Kolmogorov's theorems states

$$\lim_{n\to\infty} P\{n^{\frac{1}{2}} D_n^+ \leqq \alpha\} = 1 - e^{-2\alpha^2}, \tag{2.3}$$

and for our purposes it will be sufficient to justify Doob's method for this particular theorem since the justification of the method in general follows from it. Following Doob, define

$$x_n(t) = n^{\frac{1}{2}}\left(\frac{\nu_n(t)}{n} - t\right), \qquad 0 \leqq t \leqq 1. \tag{2.4}$$

[1] Research begun while the writer was a member of an ONR sponsored project in probability at Cornell University.

[2] For ease of comparison, we are using Doob's notation wherever possible.

Clearly,

$$(2.5)\qquad \begin{aligned} &E\{x_n(t)\} = 0, && 0 \leqq t \leqq 1,\\ &E\{[x_n(t) - x_n(s)]^2\} = (t - s)[1 - (t - s)], && 0 \leqq s \leqq t \leqq 1. \end{aligned}$$

Let $\{x(t)\}$ be a one parameter family of random variables, $0 \leqq t \leqq 1$, with the properties:

(a) for each j, if $0 \leqq t_1 < \cdots < t_j \leqq 1$, the j-variate distribution of the variables $x(t_1), x(t_2), \cdots, x(t_j)$ is Gaussian;

(b) $$E\{x(t)\} = 0, \qquad 0 \leqq t \leqq 1,$$

(2.6) $$E\{[x(t) - x(s)]^2\} = (t - s)[1 - (t - s)], \qquad 0 \leqq s \leqq t \leqq 1;$$

(c) $$P\{x(0) = 0\} = 1.$$

The $x(t)$ process can be selected so that with probability one it has continuous sample functions. Let Y be the space of these sample functions. The $x(t)$ process selected here is such that for any j, if $0 \leqq t_1 < \cdots < t_j \leqq 1$, and if $(\alpha_1, \alpha_2, \cdots, \alpha_j)$ is an arbitrary vector, we have from the central limit theorem

$$(2.7)\qquad \lim_{n\to\infty} P\{x_n(t_i) \leqq \alpha_i;\, i = 1, 2, \cdots, j\} = P\{x(t_i) \leqq \alpha_i;\, i = 1, 2, \cdots, j\}.$$

Doob's heuristic argument consisted in assuming that in calculating asymptotic $x_n(t)$ process distributions when $n \to \infty$, one could replace the $x_n(t)$ process by the $x(t)$ process. In particular, with reference to (2.3), his assumption was that

$$(2.8)\qquad \lim_{n\to\infty} P\{n^{1/2}D_n^+ \leqq \alpha\} = P\{D^+ \leqq \alpha\},$$

where $D^+ = \max_{0\leqq t\leqq 1} x(t)$. What we wish to show, therefore, is that

$$(2.9)\qquad \lim_{n\to\infty} P\left\{\operatorname*{l.u.b.}_{0\leqq t\leqq 1}\left[n^{1/2}\left(\frac{\nu_n(t)}{n} - t\right)\right] \leqq \alpha\right\} = P\{\max_{0\leqq t\leqq 1} x(t) \leqq \alpha\}.$$

Let E_n be the event that for all t in $(0, 1)$, $\nu_n(t) \leqq \alpha n^{1/2} + nt$, and let E be the event that for all t in $(0, 1)$, $x(t) \leqq \alpha$. We can write (2.9) as

$$(2.10)\qquad \lim_{n\to\infty} P\{E_n\} = P\{E\}.$$

Let E'_n be the event that for all $i = 1, 2, \cdots, n$, $\nu_n(i/n) \leqq \alpha n^{1/2} + i$, and let E''_n be the event that for all $i = 1, 2, \cdots, n$, $\nu_n(i/n) \leqq \alpha n^{1/2} + i - 1$. We have, clearly, $E''_n \subset E_n \subset E'_n$. In what follows we will show that

$$(2.11)\qquad \lim_{n\to\infty} P\{E'_n\} = P\{E\},$$

and an exactly similar argument shows $\lim_{n\to\infty} P\{E''_n\} = P\{E\}$. Hence, we will have shown (2.10).

To show (2.11), let N be a Poisson distributed random variable with mean n and independent of the random variables $x_1, x_2, x_3, \cdots$. We have, clearly,

$$P\{E'_n\} = P\{\nu_N\left(\frac{i}{n}\right) \leqq \alpha n^{\frac{1}{2}} + i; i = 1, 2, \cdots, n \mid N = n\}. \tag{2.12}$$

Let $y_1 = \nu_N(1/n)$, $y_i = \nu_N(i/n) - \nu_N((i-1)/n)$, $i = 2, 3, \cdots, n$. The variables $y_1, y_2, \cdots, y_n$ are independent (cf. Kac [4]), are Poisson distributed with mean 1, and if we let $z_i = y_i - 1$, $i = 1, 2, \cdots, n$, $s_m = z_1 + z_2 + \cdots + z_m$, then s_m is a sum of independent variables and we can rewrite (2.12) as

$$P\{E'_n\} = P\{s_i \leqq \alpha n^{\frac{1}{2}}; \quad i = 1, 2, \cdots, n \mid s_n = 0\}. \tag{2.13}$$

Now,

$$1 - P\{E'_n\} = \sum_{r=1}^{n} P\{s_i \leqq \alpha n^{\frac{1}{2}}; i = 1, 2, \cdots, r-1, s_r > \alpha n^{\frac{1}{2}} \mid s_n = 0\}. \tag{2.14}$$

Let k be a fixed positive integer; define $n_j = [jn/k]$, $j = 0, 1, 2, \cdots, k$, and let an $\epsilon > 0$ be given. From (2.14) we obtain

$$\begin{aligned} 1 - P\{E'_n\} = \sum_{j=0}^{k-1} \sum_{n_j < r \leqq n_{j+1}} & P\{s_i \leqq \alpha n^{\frac{1}{2}}; i = 1, 2, \cdots, r-1, \\ & s_r > \alpha n^{\frac{1}{2}}, |s_{n_{j+1}} - s_r| < \epsilon n^{\frac{1}{2}} \mid s_n = 0\} \\ + \sum_{j=0}^{k-1} \sum_{n_j < r \leqq n_{j+1}} & P\{s_i \leqq \alpha n^{\frac{1}{2}}; i = 1, 2, \cdots, r-1, s_r > \alpha n^{\frac{1}{2}}, \\ & |s_{n_{j+1}} - s_r| \geqq \epsilon n^{\frac{1}{2}} \mid s_n = 0\}. \end{aligned} \tag{2.15}$$

Let $P_{n,k}(\alpha) = P\{s_{n_j} \leqq \alpha n^{\frac{1}{2}}; j = 1, 2, \cdots, k \mid s_n = 0\}$. Clearly,

$$P\{E'_n\} \leqq P_{n,k}(\alpha), \tag{2.16}$$

and also the first sum on the right of (2.15) is less than $1 - P_{n,k}(\alpha - \epsilon)$. The second sum on the right of (2.15) can be written as (cf. Chung [5], pp. 39–41)

$$\begin{aligned} & \frac{n!\,e^n}{n^n} \sum_{j=0}^{k-1} \sum_{n_j < r \leqq n_{j+1}} P\{s_i \leqq \alpha n^{\frac{1}{2}}; i = 1, 2, \cdots, r-1, s_r > \alpha n^{\frac{1}{2}}, \\ & \qquad |s_{n_{j+1}} - s_r| \geqq \epsilon n^{\frac{1}{2}}, s_n = 0\} \\ & = \frac{n!\,e^n}{n^n} \Bigg[\sum_{j=0}^{k-2} \sum_{n_j < r \leqq n_{j+1}} P\{s_i \leqq \alpha n^{\frac{1}{2}}; i = 1, 2, \cdots, r-1, s_r > \alpha n^{\frac{1}{2}}\} \\ & \quad \cdot \sum_y P\{|s_{n_{j+1}} - s_r| \geqq \epsilon n^{\frac{1}{2}}, s_{n_{j+1}} = y\} P\{s_n - s_{n_{j+1}} = -y\} \\ & \quad + \sum_{n_{k-1} < r \leqq n_k} P\{s_i \leqq \alpha n^{\frac{1}{2}}; i = 1, 2, \cdots, r-1, s_r > \alpha n^{\frac{1}{2}}, \\ & \qquad |s_n - s_r| \geqq \epsilon n^{\frac{1}{2}}, s_n = 0\} \Bigg]. \end{aligned} \tag{2.17}$$

To estimate the first term in the brackets we note that since the z_i's are distributed as follows:

$$P\{z_i = m - 1\} = \frac{e^{-1}}{m!}, \qquad m = 0, 1, 2, \cdots,$$

we have, noting the maximum term of the Poisson distribution,

$$P\{s_n - s_{n_{j+1}} = -y\} \leqq A_1 k^{\frac{1}{2}} n^{-\frac{1}{2}}, \tag{2.18}$$

where A_1 is an absolute constant. Also, from Tchebycheff's inequality we get

$$\sum_y P\{\,|\, s_{n_{j+1}} - s_r \,|\, \geqq \epsilon n^{\frac{1}{2}}, s_{n_{j+1}} = y\} = P\{\,|\, s_{n_{j+1}} - s_r \,|\, \geqq \epsilon n^{\frac{1}{2}}\} \leqq \frac{1}{k\epsilon^2}. \tag{2.19}$$

The first term in the brackets on the right of (2.17) is therefore less than $A_1 k^{-\frac{1}{2}} n^{-\frac{1}{2}} \epsilon^{-2}$.

The second term in the brackets on the right of (2.17) is less than

$$\sum_{n_{k-1} < r \leqq n} \sum_{y > \alpha n^{\frac{1}{2}}} P\{s_i \leqq \alpha n^{\frac{1}{2}};\ i = 1, 2, \cdots, r-1,\ s_r = y\} P\{s_n - s_r = -y\},$$

and using similar estimates is shown to be less than $A_2 k^{-\frac{1}{2}} n^{-\frac{1}{2}}$, where A_2 is an absolute constant. Thus, we have from (2.15)

$$1 - P\{E'_n\} \leqq 1 - P_{n,k}(\alpha - \epsilon) + \frac{n!e^n}{n^n} \frac{A_3}{k^{\frac{1}{2}} n^{\frac{1}{2}} \epsilon^2}. \tag{2.20}$$

This together with (2.16) gives us

$$P_{n,k}(\alpha - \epsilon) - \frac{n!e^n}{n^n} \frac{A_3}{k^{\frac{1}{2}} n^{\frac{1}{2}} \epsilon^2} \leqq P\{E'_n\} \leqq P_{n,k}(\alpha). \tag{2.21}$$

From (2.7) we have

$$\lim_{n \to \infty} P_{n,k}(\alpha) = P\left\{x\left(\frac{i}{k}\right) \leqq \alpha,\ \ i = 1, 2, \cdots, k\right\}. \tag{2.22}$$

If in (2.21) we hold k and ϵ fixed and let $n \to \infty$, we get from (2.22) and Stirling's formula that

$$\begin{aligned} P\left\{x\left(\frac{i}{k}\right) \leqq \alpha - \epsilon;\ \ i = 1, 2, \cdots, k\right\} - \frac{\sqrt{2\pi}\, A_3}{k^{\frac{1}{2}} \epsilon^2} &\leqq \varliminf_{n \to \infty} P\{E'_n\} \\ \leqq \varlimsup_{n \to \infty} P\{E'_n\} &\leqq P\left\{x\left(\frac{i}{k}\right) \leqq \alpha;\ \ i = 1, 2, \cdots, k\right\}. \end{aligned} \tag{2.23}$$

In (2.23), if we hold ϵ fixed and let $k \to \infty$ we get from the continuity of the $x(t)$ process that

$$\begin{aligned} P\{x(t) \leqq \alpha - \epsilon,\ \ t \,\varepsilon\, (0, 1)\} \leqq \varliminf_{n \to \infty} P\{E'_n\} &\leqq \varlimsup_{n \to \infty} P\{E'_n\} \\ &\leqq P\{x(t) \leqq \alpha, t \,\varepsilon\, (0, 1)\}. \end{aligned}$$

Now finally, using the fact that the distribution function of $\max_{0 \leq t \leq 1} x(t)$ is continuous, and letting $\epsilon \to 0$ we obtain the desired statement (2.11).

3. Extension. Having shown that

$$\lim_{n \to \infty} P\{\text{l.u.b.}_{0 \leq t \leq 1}\ x_n(t) \leq \alpha\} = P\{\max_{0 \leq t \leq 1} x(t) \leq \alpha\}, \tag{3.1}$$

it is possible, using methods identical to those used by the writer in a recent paper (Donsker [6]), to obtain a general theorem like (3.1), but where the functional $\max_{0 \leq t \leq 1} x(t)$ is replaced by an arbitrary functional $F[x(t)]$ subject to certain restrictions. Indeed, we can obtain the following theorem.

THEOREM. *Let R be the space of real, single-valued functions $g(t)$ which are continuous on $0 \leq t \leq 1$ except for at most a finite number of finite jumps. Let $F[g]$ be a functional defined on R and continuous in the uniform topology at almost all points of Y*[3]*. Then,*

$$\lim_{n \to \infty} P\{F[x_n(t)] \leq \alpha\} = P\{F[x(t)] \leq \alpha\} \tag{3.2}$$

at all points of continuity of the distribution function on the right.

This theorem is proved (precisely as is the main theorem in [6]) by first obtaining (3.2) for functionals of the form $f(u_1, u_2, \cdots, u_{2k})$, where $u_i = \sup g(t)$ for $(i-1)/k < t \leq i/k$ and $u_{k+i} = \inf g(t)$ for $(i-1)/k < t \leq i/k$, $i = 1, 2, \cdots, k$, where $f(u_1, u_2, \cdots, u_{2k})$ as a function of its $2k$ variables is bounded on the whole space, Borel measurable and Riemann integrable on every finite $2k$-dimensional interval. Such a theorem is obtainable from (3.1), and moreover these functionals can be used to approximate functionals $F[g]$ which are bounded on R and continuous in the uniform topology at almost all points of Y. The approximation is such that (3.2) can be obtained for this latter class of functionals. Finally, the assumption that $F(g)$ be bounded on R may be removed, and hence we can obtain the theorem stated above, by considering the functional $e^{itF(g)}$ and using the continuity theorem for characteristic functions.

REFERENCES

[1] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[2] A. Kolmogorov, "Sulla determinazione empirica di une legge di distribuzione," *Giorn. Ist. Ital. Attuari*, Vol. 4 (1933), pp. 83–91.

[3] N. Smirnov, "Sur les écarts de la courbe de distribution empirique," *Rec. Math.* (*Mat. Sbornik*) (*NS*), Vol. 6 (1939), pp. 3–26.

[4] M. Kac, "On deviations between theoretical and empirical distribution functions," *Proc. Nat. Acad. Sci.*, Vol. 35 (1949), pp. 252–257.

[5] K. L. Chung, "An estimate concerning the Kolmogorov limit distribution," *Trans. Am. Math. Soc.*, Vol. 67 (1949), pp. 36–50.

[6] M. D. Donsker, "An invariance principle for certain probability limit theorems," *Memoirs Am. Math. Soc.*, No. 6, 1951, 12 pp.

[3] The space Y is defined above just after (2.6).

A NOTE ON THE CONVOLUTION OF UNIFORM DISTRIBUTIONS[1]

By Edwin G. Olds

Carnegie Institute of Technology

1. Introduction. Some time ago, when Dr. Acton and the present author were preparing a paper [1] on the combination of tolerances, the question arose as to the distribution of the sum of rectangular random variables having unequal bases. (For equal bases, the distribution has been known since Laplace.) As Acton pointed out, the distribution can be obtained by operational calculus. However, it seems useful to outline a derivation requiring only the well known formula for the probability density function for the sum of two random variables. In addition, this note gives several other results which may be needed in statistical quality control.

2. The distribution of the sum. Let x_i be independent random variables with probability density functions

$$f_i(x_i) = [\varepsilon(x_i) - \varepsilon(x_i - a_i)]/a_i \quad (a_i > 0; i = 1, 2, \cdots, n),$$

where $\varepsilon(x - c)$ is unity for $x \geqq c$ and zero elsewhere. Let $s = \sum_1^n x_i$ and let $f_n(s)$ and $F_n(s)$ represent the probability density function and cumulative distribution function of s respectively. Then it will be proved that

$$(1) \quad \begin{aligned} f_n(s) = \Big[s^{n-1}\varepsilon(s) &- \sum_1^n (s - a_i)^{n-1}\varepsilon(s - a_i) \\ &+ \sum_{i<j} (s - a_i - a_j)^{n-1}\varepsilon(s - a_i - a_j) - \cdots \\ &+ (-1)^n \left(s - \sum a_i\right)^{n-1}\varepsilon\left(s - \sum a_i\right)\Big] \Big/ \Big[(n-1)! \prod a_i\Big], \end{aligned}$$

and

$$(2) \quad \begin{aligned} F_n(s) = \Big[s^{n}\varepsilon(s) &- \sum_1^n (s - a_i)^{n}\varepsilon(s - a_i) \\ &+ \sum_{i<j} (s - a_i - a_j)^{n}\varepsilon(s - a_i - a_j) - \cdots \\ &+ (-1)^n \left(s - \sum a_i\right)^{n}\varepsilon\left(s - \sum a_i\right)\Big] \Big/ \Big[n! \prod a_i\Big]. \end{aligned}$$

The proof is by induction. Using the convolution formula, ([2] p. 191), we have, in our notation,

---

[1] Presented at the Annual Meeting of the Institute of Mathematical Statistics at Chicago, December 29, 1950.

Some revision of this paper was made under partial support from a U. S. Air Force contract with the Department of Mathematics of Carnegie Institute of Technology.

$$f_2(s) = \int_{-\infty}^{+\infty} f_1(s - t) f_2(t)\, dt$$

$$(3) \quad = \frac{1}{a_1 a_2}\left\{\int_{+\infty}^{+\infty} \varepsilon(t)\varepsilon(s - t)\, dt - \int_{-\infty}^{+\infty} \varepsilon(t)\varepsilon(s - t - a_1)\, dt \right.$$

$$\left. - \int_{-\infty}^{+\infty} \varepsilon(t - a_2)\varepsilon(s - t)\, dt + \int_{-\infty}^{+\infty} \varepsilon(t - a_2)\varepsilon(s - t - a_1)\, dt\right\}.$$

The first integral within the braces is zero for $t < 0$ and for $t > s > 0$ and is unity between zero and $s \geqq 0$ so the effective limits are zero and s. Likewise the effective limits for the second integral are zero and $s - a_1$. After replacing $t - a_2$ by t in the last two integrals and making obvious changes of limits, we get

$$(4) \quad f_2(s) = \frac{1}{a_1 a_2}\left\{\int_0^s \varepsilon(t)\varepsilon(s - t)\, dt - \int_0^{s-a_1} \varepsilon(t)\varepsilon(s - t - a_1)\, dt \right.$$

$$\left. - \int_0^{s-a_2} \varepsilon(t)\varepsilon(s - t - a_2)\, dt - \int_0^{s-a_1-a_2} \varepsilon(t)\varepsilon(s - t - a_1 - a_2)\, dt\right\}.$$

Finally,

$$(5) \quad f_2(s) = [s\varepsilon(s) - (s - a_1)\varepsilon(s - a_1) - (s - a_2)\varepsilon(s - a_2) + (s - a_1 - a_2)\varepsilon(s - a_1 - a_2)]/(a_1 a_2).$$

To complete the induction we need only assume that (1) holds for $n = k$ and show that it then is true for $n = k + 1$. Using the same method for combining the density functions of $s = \sum_1^k x_i$ and x_{k+1} as was used above for x_1 and x_2, this presents no difficulty. Also it is easy to show that (2) is a direct consequence of (1).

3. Asymptotic normality. For use in the remaining sections it is noted that the constants of the distribution of s are:

$$(6) \quad \text{mean, } \mu_s = \tfrac{1}{2}\sum a_i; \text{ variance, } \sigma_s^2 = \sum a_i^2/12; \text{ skewness, } \gamma_1 = 0;$$

$$\text{excess, } \gamma_2 = -\tfrac{6}{5}\sum a_i^4/(\sum a_i^2)^2.$$

The matter of convergence of the classically normed sum to the Gaussian distribution with zero mean and unit variance can be settled easily by using the well known Lindeberg condition, the sufficiency of which, as Loève [3] notes, was established by Lindeberg and by P. Lévy, and the necessity by Feller. (For discussion and references see Loève's paper.)

As the second part of the solution of the classical central limit problem Loève ([3], p. 326) states the theorem:

NC holds and $\max_{k \leq n} \dfrac{\sigma(x_k)}{\sigma(s_n)} \to 0$ *if, and only if, for every* $\epsilon > 0$

$$\sum_{k=1}^{n} \frac{1}{\sigma^2(s_n)} \int_{|x| > \epsilon\sigma(s_n)} x^2\, dF_k(x + Ex_k) \to 0.$$

For our case

$$(7) \qquad \sigma(x_k) = \frac{a_k}{\sqrt{12}}, \qquad \sigma(s_n) = \sigma_s = \sqrt{\frac{\sum_1^n a_i^2}{12}}, \qquad Ex_k = \tfrac{1}{2}a_k.$$

In order to specialize the above theorem to our use, we first establish the following:

Lemma. *The Lindeberg condition holds if, for* $k \leqq n$,

$$(8) \qquad \frac{\max a_k}{\sqrt{\sum_1^n a_i^2}} \to 0.$$

To prove this lemma it is sufficient to note, first, that each term of the Lindeberg sum is identically zero whenever $\epsilon\sigma(s_n)$ is greater than $\frac{1}{2}a_k$, and second, that the condition imposed in the lemma implies the existence of an N for any $\epsilon > 0$ such that for all $n > N$

$$(9) \qquad \frac{a_k}{\sqrt{\sum_1^n a_i^2}} < \frac{\epsilon}{\sqrt{3}}.$$

The following theorem[2] can now be established.

Theorem. *A necessary and sufficient condition for the asymptotic normality of a sum of independent rectangular random variables is*

$$(10) \qquad \lim_{\substack{n \to \infty \\ (k \leq n)}} \frac{\max a_k}{\sqrt{\sum_1^n a_i^2}} = 0.$$

To prove the sufficiency of this condition we note that by virtue of the above lemma, the Lindeberg condition is satisfied and then note, from the quoted theorem, that the Lindeberg condition implies normal convergence. For necessity, we note that if the condition fails then the Lindeberg condition must fail for, otherwise, the quoted theorem would lead to a contradiction.

Of course the condition on max a_k implies that $(a_k/\sigma_s) \to 0$ and that $\sigma_s \to \infty$. Thus any sum for which $a_i = ra_{i+1}$ will converge to normality only if $r = 1$, since, for $r > 1$, $(a_n/\sigma_s) \nrightarrow 0$, and for $r < 1$, σ_s is bounded.

4. Percentage outside three-sigma limits. For statistical quality control there is considerable interest in knowing the percentage of a distribution outside of the limits $\mu \pm 3\sigma$. For any particular sum of rectangulars this percentage can be calculated from equation [2] above. Often the total range of nonzero probability for the sum will not exceed 6σ so that the required probability will be zero.

[2] The author is grateful to the referee for suggesting a slightly different form of this theorem as an improvement of the author's original treatment, which used Lyapunov's Theorem.

It is easy to verify that this condition will hold whenever $a_{i+1} = ra_i$ and either $0 \leqq r \leqq 0.5$ or $r \geqq 2$.

When the range for the sum is greater than 6σ, an approximation to the required percentage can be obtained from an Edgeworth series. (For discussion, see [2], pp. 221–231.) Let x be the standardized variable $(s - \mu)/\sigma_s$, $\Phi(x)$ the normal distribution function, $\phi(x)$ the normal density function, and $\phi^{(i)}(x)$ its ith derivative; then, following Cramér, we have approximately

$$f(x) = \phi(x) - \frac{\gamma_1}{3!}\phi^{(3)}(x) + \frac{\gamma_2}{4!}\phi^{(4)}(x) \frac{10\gamma_1^2}{6!}\phi^{(6)}(x). \tag{11}$$

Then, integrating and substituting the pertinent values from (6) above, we have

$$F(x) = \Phi(x) - \frac{\phi^{(3)}(x)}{20}\left[\frac{\sum a_i^4}{(\sum a_i^2)^2}\right]. \tag{12}$$

Since the lower three-sigma limit for s corresponds to $x = -3$ we have, finally, for the approximate percentage below this limit

$$F(-3) = 0.00135 - 0.004\,[\textstyle\sum a_i^4/(\sum a_i^2)^2], \tag{13}$$

where the bracketed quantity takes its minimum value, n^{-1}, when all of the a's are equal.

The multiplier of the bracket has been rounded off for easy use. A better value for it is 0.0039885. Using this instead of 0.004 in (13), and making a comparison of (2) and (13) when $a_i = 1$ $(i = 1, 2, \cdots, 8)$ and $a_9 = 2$, the result from the former formula is 0.000694 and from the latter 0.000685. Using 0.004, (13) gives the approximate value 0.00068.

5. An application. The natural tolerance limits for a controlled process often are taken as $\mu \pm 3\sigma$. Let us suppose that the individual components are symmetrically distributed originally and then are symmetrically truncated by inspection with bases, a_i. Birnbaum [4] has proved that the distribution of the sum of the truncated variables is "more peaked" about the mean than the distribution of the sum of rectangular variables with the same bases, where "more peaked" means less probability for values more than any arbitrary distance from the mean. Thus, as Birnbaum points out for the case of equal truncations, the distribution of the sum of rectangulars can be used to get an upper bound for useful probabilities required for the sum of the truncated variables. For symmetric but unequal truncations, an upper bound to the percentage outside natural tolerance limits can be calculated by using formula (2) above.

REFERENCES

[1] F. S. Acton and E. G. Olds, "Tolerances—additive or Pythagorean?" *Industrial Quality Control*, Vol. 5 (1948), pp. 6–12.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] M. Loève, "Fundamental limit theorems of probability theory," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 321–338.

[4] Z. W. Birnbaum, "On random variables with comparable peakedness," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 76–80.

A LOWER BOUND FOR A PROBABILITY MOMENT OF ANY ABSOLUTELY CONTINUOUS DISTRIBUTION WITH FINITE VARIANCE

By Sigeiti Moriguti

University of North Carolina

Summary. The greatest lower bound of the nth probability moment (1.1) of a population with variance σ^2 is given by (3.4).

1. Introduction. The nth probability moment of a population with the probability density function $f(x)$ is defined as

$$\Omega_n = \int_{-\infty}^{\infty} [f(x)]^n \, dx. \tag{1.1}$$

These functionals have drawn the attention of some authors (see for instance [1] and the references given there) in connection with fitting frequency curves by means of frequency moments. Also it is to be noted[1] that the cumulative distribution function of the range w of a sample of size n from such a population can be approximated, for small w, by $n\Omega_n w^{n-1}$.

In general, it is not necessary that n in (1.1) be an integer. It may be any real number. However we put the restriction $n > 1$ in the following. To be specific, we take the population mean equal to zero. Moreover, we consider only populations whose variance

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) \, dx \tag{1.2}$$

is finite.

As the probability density function, $f(x)$ must satisfy the conditions

$$\int_{-\infty}^{\infty} f(x) \, dx = 1, \tag{1.3}$$

$$f(x) \geqq 0. \tag{1.4}$$

Under these conditions, we try to find a lower bound for Ω_n.

Incidentally, Ω_n has no finite upper bound, because it increases indefinitely as, for instance, the probability concentrates more and more to a certain point.

2. Derivation of the extremal distribution. The calculus of variations suggests equating to zero the first variation

$$\delta \left[\int_{-\infty}^{\infty} [f(x)]^n \, dx - \lambda \int_{-\infty}^{\infty} x^2 f(x) \, dx - \mu \int_{-\infty}^{\infty} f(x) \, dx \right], \tag{2.1}$$

[1] The author in fact took up this problem at first in connection with his work on the distribution of sample ranges. It was Professor Harold Hotelling who called the author's attention to probability moments in this relation.

where λ and μ are the Lagrange multipliers. Thus we get as the characteristic equation

$$n[f(x)]^{n-1} - \lambda x^2 - \mu = 0, \tag{2.2}$$

whence

$$f(x) = \left[\frac{\lambda x^2 + \mu}{n}\right]^{1/(n-1)}. \tag{2.3}$$

We should take λ negative, and consequently μ positive. Then the solution (2.3) is applicable in the interval $(-\sqrt{-\mu/\lambda},\ \sqrt{-\mu/\lambda})$. Outside of the interval, $f(x)$ should be taken to be identically equal to zero.

TABLE 1

Reduced probability moment, $\Omega_n \sigma^{n-1}$

Order	Lower bound	Normal distribution	Rectangular distribution	Asymptotic formula
2	.26833	.28209	.28868	.2547
3	.07599	.09189	.08333	.0735
4	.02174	.03175	.02406	.0212
5	$.0^2 6245$	.01133	$.0^2 6944$	$.0^2 613$
6	$.0^2 1797$	$.0^2 4125$	$.0^2 2005$	$.0^2 177$
7	$.0^3 5174$	$.0^2 1524$	$.0^3 5787$	$.0^3 511$
8	$.0^3 1491$	$.0^3 5686$	$.0^3 1671$	$.0^3 147$
9	$.0^4 4299$	$.0^3 2139$	$.0^4 4823$	$.0^4 426$
10	$.0^4 1240$	$.0^4 8094$	$.0^4 1392$	$.0^4 123$

As a change of scale in measuring x does not affect the result essentially, we take the nonvanishing interval to be $(-1, 1)$, and write the solution in the form

$$\begin{aligned} f(x) &= c(1 - x^2)^{1/(n-1)}, & -1 \leqq x \leqq 1, \\ &= 0, & |x| > 1, \end{aligned} \tag{2.4}$$

where c is determined by the normalizing condition (1.3) as

$$c = \frac{1}{\mathrm{B}\left(\frac{n}{n-1}, \frac{1}{2}\right)}. \tag{2.5}$$

The variance of the distribution (2.4) is calculated as

$$\sigma^2 = \frac{n-1}{3n-1}. \tag{2.6}$$

The nth probability moment of the distribution (2.4) is

$$\Omega_n = \frac{2n}{3n-1} c^{n-1}. \tag{2.7}$$

Therefore the reduced nth probability moment $\Omega_n\sigma^{n-1}$, which is invariant under any linear transformation of x, is given for the distribution of the same type as (2.4) by

$$\Omega_n\sigma^{n-1} = \frac{2n}{3n-1}\left[\sqrt{\frac{n-1}{3n-1}}\Big/ \mathrm{B}\left(\frac{n}{n-1}, \frac{1}{2}\right)\right]^{n-1}. \tag{2.8}$$

That this value gives the lower bound for any population with finite variance is to be proved in the next section.

3. Proof that the solution gives the lower bound. Let us denote the particular probability density function (2.4) by $\bar{f}(x)$, and compare the probability moment $\bar{\Omega}_n$ for it with Ω_n for any distribution with probability density function $f(x)$ and the same variance σ^2. From the normalizing condition and the assumed equality of the variance, we get

$$\int_{-\infty}^{\infty} [f(x) - \bar{f}(x)]\, dx = 0, \quad \int_{-\infty}^{\infty} x^2[f(x) - \bar{f}(x)]\, dx = 0. \tag{3.1}$$

By virtue of these equations and taking account of (2.4), we can express the difference $\Omega_n - \bar{\Omega}_n$ in the following way:

$$\begin{aligned} \Omega_n - \bar{\Omega}_n &= \int_{-\infty}^{\infty} [\{f(x)\}^n - \{\bar{f}(x)\}^n - nc^{n-1}(1 - x^2)\{f(x) - \bar{f}(x)\}]\, dx \\ &= \int_{-1}^{1} [\{f(x)\}^n - \{\bar{f}(x)\}^n - n\{\bar{f}(x)\}^{n-1}\{f(x) - \bar{f}(x)\}]\, dx \\ &\quad + \int_{|x|>1} [\{f(x)\}^n + nc^{n-1}(x^2 - 1)f(x)]\, dx. \end{aligned} \tag{3.2}$$

But Taylor's expansion up to the second-order term provides the formula, for any f and $\bar{f}$,

$$f^n = \bar{f}^n + n\bar{f}^{n-1}(f - \bar{f}) + \frac{n(n-1)}{2} f_1^{n-2}(f - \bar{f})^2, \tag{3.3}$$

where f_1 is a value between f and $\bar{f}$. As both $f(x)$ and $\bar{f}(x)$ are positive, the formula (3.3) assures us that the integrand of the first integral in the last member of (3.2) is nonnegative. Also the integrand of the second integral is obviously nonnegative. Hence we get the conclusion $\Omega_n \geqq \bar{\Omega}_n$, equality being satisfied only if $f(x) \equiv \bar{f}(x)$.

In general, it is easily derived from the above and (2.8) that

$$\Omega_n \geqq \frac{2n}{3n-1}\left[\sqrt{\frac{n-1}{3n-1}}\Big/ \mathrm{B}\left(\frac{n}{n-1}, \frac{1}{2}\right)\right]^{n-1} \frac{1}{\sigma^{n-1}}. \tag{3.4}$$

Thus the lower bound of a probability moment of any absolutely continuous distribution with finite variance σ^2 is given by the right-hand member of (3.4). It is actually achieved by a distribution of the same type as (2.4).

4. Numerical results. Numerical values of the coefficient in (3.4) are tabulated in Table 1, together with the corresponding values $n^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}$ for normal and $(2\sqrt{3})^{-(n-1)}$ for rectangular population. All these values approach 1 when $n \to 1$, as might be expected from the fact that for any distribution $\Omega_1 = 1$. It is to be noted that the curve for the lower bound would be fairly parallel in logarithmic scale to the curve for rectangular population. In fact it is easily shown that when n becomes large the former is given by

$$\begin{aligned}(4.1)\quad & \frac{1}{(2\sqrt{3})^{n-1}}\frac{2}{3}\exp\left[-\frac{1}{3}+\psi\left(\frac{1}{2}\right)-\psi(0)\right]\left(1+o\left(\frac{1}{n}\right)\right)\\ &= \frac{1}{(2\sqrt{3})^{n-1}}\frac{e^{5/3}}{6}\left(1+o\left(\frac{1}{n}\right)\right)=\frac{0.8824}{(2\sqrt{3})^{n-1}}\left(1+o\left(\frac{1}{n}\right)\right),\end{aligned}$$

where $\psi(x)$ denotes the digamma function $\Gamma'(x + 1)/\Gamma(x + 1)$. The first term happens to be close to the true value even for small n as we see in Table 1.

5. Acknowledgement. The author wishes to express his thanks to Professor Harold Hotelling for his kind supervision of the work.

REFERENCE

[1] HERBERT S. SICHEL, "The method of frequency-moments and its application to type VII populations," *Biometrika*, Vol. 36 (1949), pp. 404–425.

UNIFORMITY FIELD TRIALS WHEN DIFFERENCES IN FERTILITY LEVELS OF SUBPLOTS ARE NOT INCLUDED IN EXPERIMENTAL ERROR

BY G. A. BAKER

University of California, Davis

1. Introduction. The present note is confined to the consideration of two randomized blocks with two subplots each. The usual mathematical model for the analysis of variance of such an experiment assumes that

$$(1.1)\qquad v_{ij} = g + b_i + t_j + \epsilon_{ij}, \qquad i = 1, 2; j = 1, 2,$$

where v_{ij} is the yield of the jth variety in the ith block, and the block effect b_i is the average for the subplots of the ith block. Any difference between b_i and the yield of subplots due to differences in fertility is one component of the random parts, ϵ_{ij}. The random parts, ϵ_{ij}'s, are then assumed to be normally and independently distributed with *zero* means and uniform variance. That these assumptions may break down in many cases because of the magnitude and nonrandomness of the differences between subplots has been indicated in a recent paper [1]. It should be understood that it is practically impossible with our present knowledge to determine the relative or absolute fertility levels of any set of plots,

so that the present discussion will add only to the background knowledge and general understanding of the behavior of field trials. It is possible to discuss the present simple case in some detail while the situation becomes more complex with more degrees of freedom. (See [1], pages 64 and 65.) The effect of randomization is considered, and it is found to have a "beneficial" effect in some cases, and no effect in others.

2. Theoretical development. The following development follows closely suggestions made by a referee, especially with respect to randomization. Let us consider the following setup:

Block 1	Block 2
v_{11}	v_{21}
v_{12}	v_{22}

The v_{ij}'s refer to observed yields of a uniformity trial. The subscript i refers to the block number and j to a dummy variety. Since this is a uniformity trial the t_j's of equation (1.1) are zero. The plots we shall consider as being assigned at random to the dummy varieties.

Let σ_0^2 be the assumed uniform error variance and $\xi_{ih}(i = 1, 2; h = 1, 2)$ be the "true" unknown fertility level in the hth subplot of the ith block. Let

$$v_{ij} = x_{ij} + \xi'_{ij}. \tag{2.1}$$

Then it is assumed that the x_{ij}'s are distributed as $N(0, \sigma_0)$. The jth "variety" has equal chance of being assigned either to the first or to the second plot within the block. Thus ξ'_{ij} itself is a stochastic variate, with probability 1/2 of taking the values ξ_{i1}, ξ_{i2}. Put

$$b_i = \frac{\xi_{i1} + \xi_{i2}}{2}, \; d_i = \frac{\xi_{i1} - \xi_{i2}}{2} \qquad (i = 1, 2). \tag{2.2}$$

Since if variety 1 is assigned to a given plot in the ith block then variety 2 must be assigned to the second plot we have

$$\begin{aligned} v_{11} &= b_1 + x_{11} + a_1, & v_{21} &= b_2 + x_{21} + a_2, \\ v_{12} &= b_1 + x_{12} - a_1, & v_{22} &= b_2 + x_{22} - a_2, \end{aligned} \tag{2.3}$$

where a_i is a stochastic variate which takes the values $\pm d_i$ with equal probabilities 1/2.

If we apply the conventional analysis of variance we obtain

$$S_v^2 = \tfrac{1}{4}(v_{11} + v_{21} - v_{12} - v_{22})^2, \tag{2.4}$$

$$S_e^2 = \tfrac{1}{4}(v_{11} - v_{21} - v_{12} + v_{22})^2, \tag{2.5}$$

where S_v^2 is the variety sum of squares and S_e^2 is the error sum of squares, each with one degree of freedom. These expressions (2.4) and (2.5) in terms of x_{ij}'s and a_i's are

$$S_v^2 = \tfrac{1}{4}(x_{11} + x_{21} - x_{12} - x_{22} + 2a_1 + 2a_2)^2, \tag{2.6}$$

$$S_e^2 = \tfrac{1}{4}(x_{11} - x_{21} - x_{12} + x_{22} + 2a_1 - 2a_2)^2. \tag{2.7}$$

Put

$$\begin{aligned} v &= x_{11} + x_{21} - x_{12} - x_{22}, \\ u &= x_{11} - x_{21} - x_{12} + x_{22}, \\ m_2' &= 2a_1 + 2a_2, \\ m_1' &= 2a_1 - 2a_2, \end{aligned} \tag{2.8}$$

and we get

$$F_r = \left(\frac{v + m_2'}{u + m_1'}\right)^2, \tag{2.9}$$

where u and v are independent variates distributed $N(0, 4\sigma_0^2)$. Put

$$\begin{aligned} m_1 &= \xi_{11} - \xi_{12} - \xi_{21} + \xi_{22}, \\ m_2 &= \xi_{11} - \xi_{12} + \xi_{21} - \xi_{22}, \end{aligned} \tag{2.10}$$

and then the pair (m_2', m_1') has the four possible values (m_2, m_1), (m_1, m_2), $(-m_2, -m_1)$, $(-m_1, -m_2)$, each with probability 1/4.

If we used a systematic arrangement with the variety number the same as subplot number instead of a randomized arrangement we would have

$$F = \left(\frac{v + m_2}{u + m_1}\right)^2 \tag{2.11}$$

instead of (2.9). If we apply the result given in [2], especially equation (17) page 5, and transform to a new variable, we obtain the distribution of F as

$$\begin{aligned} f(F, m_2, m_1) &= \pi^{-1}(1 + F)^{-1}F^{-\frac{1}{2}}e^{-\frac{1}{2}(m^2{}_1+m_2{}^2)/\sigma^2} \\ &+ a(2\pi)^{-\frac{1}{2}}(1 + F)^{-1}F^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(m_1F^{\frac{1}{2}} - m_2)^2/(\sigma^2(1 + F))\right] \int_0^a N(x)\, dx \\ &+ b(2\pi^{-\frac{1}{2}})(1 + F)^{-1}F^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(m_2 + m_1F^{\frac{1}{2}})^2/(\sigma^2(1 + F))\right] \int_0^b N(x)\, dx, \end{aligned} \tag{2.12}$$

$0 \leqq F \leqq \infty$, where

$$\begin{aligned} a &= (m_2F^{\frac{1}{2}} + m_1)/(\sigma(1 + F)^{\frac{1}{2}}), \\ b &= (m_1 - m_2F^{\frac{1}{2}})/(\sigma(1 + F)^{\frac{1}{2}}), \\ N(x) &= (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}. \end{aligned}$$

We note that if both m_1 and m_2 are zero (2.12) reduces to the tabled F distribution which is used almost universally in testing the significance of "variety" difference with 1 and 1 degrees of freedom.

TABLE 1

Distributions (2.12) and (2.13) for seven pairs of values of the parameters

	DESIGNATING NUMBER OF PARAMETRIC PAIR								
F	1[c]	2[a]	3[a]	$\frac{1}{2}(2+3)$[b]	4[c]	5[a]	6[a]	$\frac{1}{2}(5+6)$[b]	7[c]
0.0001	31.828	19.307	46.529	32.918	28.229	4.309	80.439	42.374	10.893
0.005	4.479	2.724	6.524	4.624	3.992	0.618	11.185	5.902	1.620
0.010	3.152	1.948	4.572	3.260	2.822	0.444	7.774	4.109	1.187
0.025	1.964	1.221	2.817	2.019	1.783	0.293	4.674	2.483	0.862
0.05	1.356	0.862	1.909	1.385	1.257	0.221	3.045	1.633	0.712
0.10	0.915	0.607	1.244	0.926	0.879	0.175	1.844	1.010	0.650
0.20	0.593	0.423	0.758	0.591	0.613	0.147	0.987	0.567	0.604
0.40	0.360	0.286	0.417	0.352	0.388	0.131	0.440	0.285	0.518
0.60	0.267	0.238	0.276	0.257	0.282	0.126	0.249	0.188	0.427
0.80	0.198	0.183	0.200	0.192	0.220	0.116	0.158	0.137	0.348
1.00	0.159	0.154	0.154	0.154	0.177	0.109	0.109	0.109	0.284
1.20	0.132	0.132	0.123	0.127	0.147	0.103	0.080	0.091	0.233
1.40	0.112	0.116	0.101	0.108	0.124	0.097	0.060	0.078	0.194
1.60	0.096	0.103	0.084	0.094	0.107	0.091	0.048	0.069	0.162
1.80	0.085	0.092	0.072	0.082	0.093	0.086	0.038	0.062	0.138
2.00	0.075	0.084	0.062	0.073	0.082	0.081	0.032	0.056	0.118
2.20	0.067	0.076	0.055	0.065	0.073	0.076	0.027	0.052	0.102
2.40	0.060	0.070	0.049	0.059	0.065	0.072	0.023	0.048	0.089
2.60	0.055	0.064	0.043	0.054	0.059	0.068	0.020	0.044	0.078
2.80	0.050	0.059	0.039	0.049	0.053	0.065	0.017	0.041	0.068
3.00	0.046	0.055	0.035	0.045	0.048	0.062	0.015	0.038	0.061
3.20	0.042	0.051	0.032	0.042	0.043	0.059	0.013	0.036	0.054
3.40	0.039	0.048	0.030	0.039	0.041	0.056	0.012	0.034	0.049
3.60	0.036	0.045	0.027	0.036	0.038	0.053	0.011	0.032	0.044
3.80	0.034	0.042	0.025	0.034	0.035	0.051	0.010	0.030	0.040
4.00	0.032	0.040	0.024	0.032	0.033	0.049	0.009	0.029	0.036
4.20	0.030	0.037	0.022	0.030	0.031	0.046	0.008	0.027	0.033
4.40	0.028	0.035	0.020	0.028	0.029	0.045	0.007	0.026	0.031
4.60	0.026	0.034	0.019	0.026	0.027	0.043	0.007	0.025	0.028
4.80	0.025	0.032	0.018	0.025	0.025	0.041	0.006	0.024	0.026
5.00	0.024	0.030	0.017	0.024	0.024	0.040	0.006	0.023	0.024
100	0.(2)03[d]	0.(2)05	0.(3)19	0.(3)34	0.(3)14	0.(3)77	0.(3)04	0.(2)04	0.(3)11
10,000	0.(6)32	0.(6)46	0.(6)19	0.(6)33	0.(6)28	0.(6)84	0.(7)43	0.(6)44	0.(6)11
Limiting ratio[e]	1.	1.462	0.606	1.034	0.887	1.332	0.135	0.733	0.180

[a] Distributions headed 2, 3, 5, and 6 are not randomized (2.12).
[b] Distributions $(2+3)/2$, $(5+6)/2$ are randomized (2.13).
[c] Distributions 1, 4, and 7 are identical under randomization and nonrandomization.
[d] Number in parenthesis indicates the number of omitted zeros, thus 0.(2)03 means 0.0003.
[e] Limiting ratio of the ordinates of the indicated distributions to the ordinates of the conventional F distribution (1) as F approaches zero and as F approaches infinity.

The distribution of F when randomization is allowed is

$$\tag{2.13} \tfrac{1}{2}[f(F, m_2, m_1) + f(F, m_1, m_2)],$$

since $f(F, m_2, m_1) = f(F, -m_2, -m_1)$.

3. Discussion. To show how (2.12) and (2.13) may differ from the usually assumed distribution we have considered the following seven pairs of values of m_1/σ and m_2/σ:

DESIGNATING NUMBER	$\frac{m_1}{\sigma}$	$\frac{m_2}{\sigma}$
1	0	0
2	0	1
3	1	0
4	1	1
5	0	2
6	2	0
7	2	2

Selected ordinates for systematic and randomized procedures for these 7 pairs of values are presented and compared in Table 1. It is seen that the tails of some of the curves are much heavier than for case 1 ($m_1 = m_2 = 0$), indicating that much larger values of F are required for significance. On the other hand, some of the tails are lighter than for case 1 so that smaller F-values are indicative of significance at the usual levels. Randomization is effective in some cases in giving a distribution that is closer to the conventional F distribution than is the F distribution for a systematic procedure.

It is easy to find the limiting values of the ratios of the ordinates of (2.12) and (2.13) to the ordinates of the conventional F distribution as F approaches 0 and ∞ (same). These limiting values are also indicated in Table 1.

When (2.13) is a greatly curtailed distribution making errors of the first kind less probable than expected then the probability of errors of the second kind may be greatly enhanced.

REFERENCES

[1] G. A. Baker and F. N. Briggs, "Yield trials with backcross derived lines of wheat," *Annals of Inst. of Stat. Math., Tokyo*, Vol. 2 (1950), pp. 61–67.

[2] G. A. Baker, "Distribution of the means divided by the standard deviations of samples from nonhomogeneous populations," *Annals of Math. Stat.*, Vol. 3 (1932), pp. 1–9.

A GENERALIZATION OF A THEOREM DUE TO MacNEISH[1]

By K. A. Bush

Champlain College, State University of New York, and University of North Carolina

1. Summary and introduction. In 1922 MacNeish [1] considered the problem of orthogonal Latin squares and showed that if the number s is written in standard form:

$$s = p_0^{n_0} p_1^{n_1} \cdots p_k^{n_k},$$

[1] This note is a revision of one section of the author's doctoral dissertation submitted to the University of North Carolina at Chapel Hill.

where $p_0, p_1, \cdots, p_k$ are primes, and if

$$r = \min(p_0^{n_0}, p_1^{n_1}, \cdots, p_k^{n_k}),$$

then we can construct $r - 1$ orthogonal Latin squares of side s. An alternative proof was also given by Mann [2]. At the April, 1950 meeting of the Institute of Mathematical Statistics at Chapel Hill, North Carolina, R. C. Bose announced an interesting generalization of this result [3] which is stated as a theorem in the next section. The proof given here is simpler than Bose's original proof and is published at his suggestion.

2. Bose's generalization of MacNeish's theorem. Let us consider a matrix $A = (a_{ij})$, where each a_{ij} represents one of the integers $0, 1, \cdots, s - 1$ with N columns and k rows. Consider all t-rowed submatrices of N columns which can be formed from this array by choosing any t rows. Each column of the submatrix can be regarded as an ordered t-plet. The matrix A will be called an orthogonal array (N, k, s, t) of size N, k constraints, s levels, strength t and index λ if each of the C_t^k t-rowed submatrices that may be formed from A contains every one of the s^t possible ordered t-plets each repeated λ times. It is clear that we cannot add rows indefinitely to the array and still preserve its orthogonal character. We shall use the symbol $f(N, s, t)$ to denote the maximum number of constraints that are possible.

THEOREM. *If N_i is divisible by s_i^t for $i = 1, 2, \cdots, u$, then*

$$f(N_1N_2 \cdots N_u, s_1s_2 \cdots s_u, t) \geqq \min(k_1, k_2, \cdots, k_u),$$

where $k_i = f(N_i, s_i, t)$.

PROOF. Let $N_i = \lambda_i s_i^t$. We shall proceed inductively, and we first establish the relationship:

$$f(N_1N_2, s_1s_2, t) \geqq \min(k_1, k_2).$$

Let us denote the orthogonal array with N_1 columns and k_1 constraints by $A = (a_{ij})$ and the second array with N_2 columns and k_2 constraints by $B = (b_{ij})$. We may regard the elements of these two arrays as elements of two additive Abelian groups. Accordingly we may form the direct sum of these two groups. There are s_1s_2 elements in this sum, and we may represent any element of this new group by the symbol (a_{ij}, b_{mn}) where a_{ij} and b_{mn} are elements of the two modules. We now write the array with N_1N_2 columns in the form

$$\begin{array}{c}
(a_{k1}, b_{k1}) \cdots (a_{kN_1}, b_{k1}) \cdots (a_{k1}, b_{kN_2}) \cdots (a_{kN_1}, b_{kN_2}) \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
(a_{11}, b_{11}) \cdots (a_{1N_1}, b_{11}) \cdots (a_{11}, b_{1N_2}) \cdots (a_{1N_1}, b_{1N_2}),
\end{array}$$

where the elements of A are used for the first component for the first N_1 columns and for the first k rows, where $k = \min(k_1, k_2)$. The construction is completed in a similar manner for the next group of N_1 columns (not indicated in the array above) and so on until N_2 groups of N_1 columns have been written down so that

$N = N_1N_2$. On the other hand, the second component is taken direct'yfrom the array $B = (b_{ij})$.

Now select any t rows from the array so constructed. Any t-plet of the b elements is repeated N_2 times in each of λ_2 groups. Within each of these groups of N_1 objects any particular t-plet of the a elements occurs λ_1 times so that each t-plet which is constructed from the compound elements occurs $\lambda_1\lambda_2$ times. Thus the new array is orthogonal.

We now adjoin the array (N_3, k_3, s_3, t), where $k = \min(k_1, k_2, k_3)$, to the one we have just constructed, by an analogous process. Continuing in this manner, we reach our theorem. In particular if $t = 2$, and $\lambda_i = 1$ for $i = 1, 2, \cdots, u$, we secure the MacNeish theorem (cf. [1]).

As an example of the use of our theorem, we can state as an illustrative result

$$f(72, 6, 2) \geqq 4$$

since $f(3^2, 3, 2) = 4, f(2^3, 2, 2) = 7$ in accordance with results established in [4]. In the absence of this extension of the MacNeish result, it might have been supposed that there could be but three orthogonal rows for this case, since there are no orthogonal Latin squares of side 6. We cannot, however, conclude that the equality sign holds since counter examples have been given in [4].

REFERENCES

[1] H. F. MacNeish, "Euler's squares," *Annals of Mathematics*, Vol. 23 (1922), pp. 221–227.

[2] H. B. Mann, "The construction of orthogonal Latin squares," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 418–423.

[3] R. C. Bose, "A note on orthogonal arrays," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 304–305 (Abstract).

[4] R. C. Bose and K. A. Bush, "Orthogonal arrays of strength two and three," *Annals of Math. Stat.*, Vol. 22 (1951), p. 312 (Abstract).

ON A LIMITING CASE FOR THE DISTRIBUTION OF EXCEEDANCES, WITH AN APPLICATION TO LIFE-TESTING

By Lee B. Harris

General Electric Company

According to equation (4.12) of [1], the probability that in a future sample of N observations, taken from an unknown distribution of a continuous variate, less than x of them will exceed x_m, the mth highest observation in the trial sample of n observations, is given by

$$W(n, m, N, x) = 1 - \frac{\binom{N}{x+1}}{\binom{N+n}{x+1}} F_m(x+1, -n, -n-N+x+1, 1),$$

where F_m is the sum of the first m terms of the hypergeometric series, having the parameters indicated in the parentheses. If we set $m = 1$, we find that the probability of getting in a future sample of N trials at most x exceedances of the largest value in a trial sample of n observations is

$$W(n, 1, N, x) = 1 - \left[\binom{N}{x+1}\bigg/\binom{N+n}{x+1}\right], \tag{1}$$

since $F_1 = 1$.

If x and N are both large, we can approximate the factorials in (1) with Stirling's formula, $a! \approx \sqrt{2\pi a}(a/e)^a$. Then (1) reduces to

$$\begin{aligned} 1 - W(n, 1, N, x) &\approx \left(1 - \frac{x+1}{N}\right)^n \\ &\cdot \left\{\frac{\left(1 + \dfrac{n}{N-x-1}\right)^{N-x-1+n}}{\left(1 + \dfrac{n}{N}\right)^{n+N}}\right\} \sqrt{1 - \frac{n}{n+N}}\sqrt{1 + \frac{n}{N-x-1}}. \end{aligned} \tag{2}$$

Now consider the limiting case in which N and x both approach infinity in such a way that $x = kN$. This is the case in which we wish to find the probability that in a very large future sample at most a fraction k of the observations will exceed the largest value in the trial sample of n observations. Considering each of the factors on the right side of (2), we have

$$\lim_{x=kN\to\infty}\left(1 - \frac{x+1}{N}\right)^n = (1-k)^n,$$

$$\lim_{x=kN\to\infty}\left(1 + \frac{n}{N-x-1}\right)^{N-x-1+n} = \lim_{x=kN\to\infty}\left(1 + \frac{n}{N}\right)^{N+n} = e^n,$$

$$\lim_{x=kN\to\infty}\sqrt{1 - \frac{n}{N+n}} = \lim_{x=kN\to\infty}\sqrt{1 + \frac{n}{N-x-1}} = 1.$$

Hence,

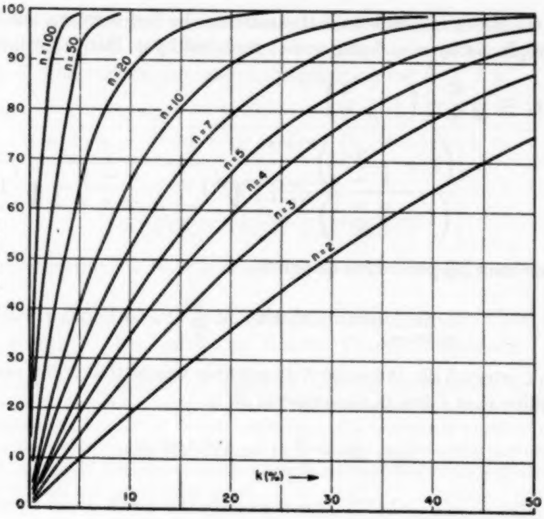
$$\lim_{N\to\infty} W(n, 1, N, kN) = 1 - (1-k)^n. \tag{3}$$

The probability density, which may be obtained from (3) by differentiation, is

$$p(k) = n(1-k)^{n-1}. \tag{4}$$

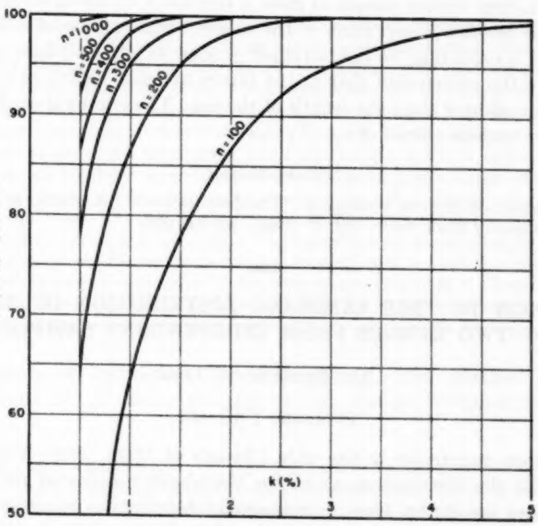
An interesting check on the consistency of the theory is a proof of (4) based on Gumbel's original discrete distribution of x. Setting $m = 1$ in equation (1.3) of [1] we have

$$w(n, 1, N, x) = \frac{n}{N+n}\left\{\frac{\dbinom{N}{x}}{\dbinom{N+(n-1)}{x}}\right\}.$$



n=100
n=50
n=20
n=10
n=7
n=5
n=4
n=3
n=2
$1-(1-k)^n$ (%)
k(%)
0
10
20
30
40
50
60
70
80
90
100
10
20
30
40
50

FIG. 1



n=1000
n=500
n=400
n=300
n=200
n=100
$1-(1-k)^n$ (%)
k(%)
50
60
70
80
90
100
1
2
3
4
5

FIG. 2

Note that the factor in brackets is the same as the last term on the right side of (1) with n replaced by $n - 1$ and $x + 1$ replaced by x. Hence for large x and N,

$$w(n, 1, N, x) \approx \frac{n}{N+n}\left(1 - \frac{x}{N}\right)^{n-1}$$

$$\cdot \left\{\frac{\left(1 + \dfrac{n}{N-x}\right)^{N-x+n}}{\left(1 + \dfrac{n-1}{N}\right)^{N+n-1}}\right\} \sqrt{1 - \frac{n-1}{N+n-1}}\, \sqrt{1 + \frac{n-1}{N-x}}.$$

By the same limiting procedure as before,

$$\lim_{x=kN\to\infty} w(n, 1, N, kN) = \frac{n}{N}(1-k)^{n-1}. \tag{5}$$

In any small interval dk, there are Ndk possible values that x can assume; hence the probability that k lies in the interval dk is

$$p(k)\, dk = \frac{n}{N}(1-k)^{n-1}(N\, dk). \tag{6}$$

Therefore, $p(k) = n(1 - k)^{n-1}$. This is exactly the result given by equation (4), but obtained in a somewhat different way.

From the symmetry of the problem, $\lim_{N\to\infty} W(n, 1, N, kN)$ is also the probability that in a large future sample at most a fraction k of the observations will be *less* than the *smallest* observation in the original trial sample of n units. Hence, a life-test of n units may be discontinued as soon as any unit fails and equation (3) will give the probability that in the future at most $100k\%$ of the units will fail in a time shorter than the length of the test. The graphs show W as a function of k for various values of n.

REFERENCE

[1] E. J. Gumbel and H. Von Schelling, "The distribution of the number of exceedances," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 247-262.

CORRECTION TO "THE SAMPLING DISTRIBUTION OF THE RATIO OF TWO RANGES FROM INDEPENDENT SAMPLES"

By Richard F. Link

Princeton University

In the note mentioned in the title (*Annals of Math. Stat.*, Vol. 21 (1950), pp. 112–116) the distribution given for the above mentioned ratio when the sample values are drawn from a rectangular distribution is correct only when $R \leqq 1$. This is pointed out in an article by P. R. Rider ("The distribution of the

quotient of ranges in samples from a rectangular population," *Jour. Am. Stat. Assn.*, Vol. 46 (1951), pp. 502–507), who also gives the correct density of the ratio for $R \geqq 1$. The correct cumulative distribution for $R \geqq 1$ is

$$1 - R^{-n_2}\left\{\frac{Rn_2n_1(n_1 - 1)}{(n_1 + n_2 - 1)(n_1 + n_2 - 2)} - \frac{n_1(n_1 - 1)(n_2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)}\right\}.$$

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Blacksburg meeting of the Institute, March 19–21, 1952)

1. On the Approximation of Sampling Distributions by Punch Card Methods. CARL F. KOSSACK AND LESTER L. HELMS, Purdue University.

This paper presents a procedure for obtaining empirical distributions, by punch card methods, of statistics for which the exact distribution or a usable approximation has not been found. The mechanization of random sampling of a univariate population has been described and extended to random sampling of a correlated multivariate population whose covariance matrix is given. This procedure has been applied to Wald's classification statistic in the univariate case, and the results noted.

2. Resolvable Incomplete Block Designs with Two Replications. R. C. BOSE AND K. R. NAIR, University of North Carolina.

Incomplete block designs in which the blocks can be grouped in such a way that each group contains a complete replication may be called resolvable designs. They are useful from the point of view of recovery of inter-block information. It is therefore important to investigate resolvable designs involving a few replications. In this paper we consider a class of resolvable designs with two replications, which contains as a special case the well known square and rectangular lattices with two replications. Given a symmetrical balanced incomplete block design with u treatments, and r replications in which each pair occurs λ times, we can use the incidence matrix (n_{ij}) of this design to form a design of one class in the following way. Take a $u \times u$ square scheme, and in the cell (i, j) put x new treatments when $n_{ij} = 1$, and y new treatments when $n_{ij} = 0$. The total number of treatments obtained in this way is $v = u[rx + (u - r)y]$. The design is now constructed by taking the rows of the scheme for the blocks of the first replication, and the columns of the scheme for the blocks of the second replication. It has been shown that both the intra- and inter-block analysis can be carried out in a simple manner. The necessary formulae have been given, and the computational procedure illustrated by working out a numerical example.

3. Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons. R. A. BRADLEY AND M. E. TERRY, Virginia Polytechnic Institute.

True preferences or ratings $\pi_{1u}, \cdots, \pi_{tu}$, $\sum_{i=1}^{t} \pi_{iu} = 1$, are assumed to exist for t treatments in the uth of g groups of experimental data in an experiment involving paired comparisons. For the uth group, the probability that treatment i is "better" than treatment j when they appear in a pair is postulated to be $\pi_{iu}/(\pi_{iu} + \pi_{ju})$.

Three tests of hypotheses are available and estimates of the treatment ratings may be

obtained. The tests use likelihood ratio statistics to test (a) $H_0 : \pi_{iu} \equiv 1/t$, against $H_1 : \pi_{iu} \equiv \pi_i$ for all u; (b) $H_0 : \pi_{iu} \equiv 1/t$, against $H_1 : \pi_{iu} \not\equiv 1/t$; and (c) $H_0 : \pi_{iu} \equiv \pi_i$ for all u, against $H_1 : \pi_{iu} \not\equiv \pi_i$.

Small-sample distributions with tables are available for tests (a) and (b). In all three tests limiting distributions are shown to be in the form of chi-square.

4. **Multiple Regression with a Quantal Response.** D. B. Duncan and R. C. Rhodes, Virginia Polytechnic Institute.

The problem considered is that of fitting a maximum likelihood multiple regression equation to data in which the response is quantal, the probit transformation is appropriate and the number, r, of independent regression variates is not small.

Iterative methods, for example the Bliss-Fisher method, are available, but these have been developed mainly for the case $r = 1$ and rapidly become impractical for cases $r > 2$.

A method is developed based on (i) the approximation of the weighted deviations of the working probits from the provisional probits by linear functions of the provisional probits and (ii) the replacement of the independent x variates throughout most of the procedure by a linear function of them, termed a composite regression variate. These devices lead to a simple procedure and result in an estimated 70 to 90% saving in work.

5. **Rank Analysis of Incomplete Block Designs. II. The Method for Blocks of Three. (Preliminary Report.)** R. A. Bradley and M. E. Terry, Virginia Polytechnic Institute.

The extensions of "Rank analysis of incomplete block designs. I. The method of paired comparisons," Abstract No. 3 above, to blocks of size three are presented. As before, true preferences or ratings $\pi_{1u}, \cdots, \pi_{tu}$, $\sum_{i=1}^{t} \pi_{iu} = 1$ are assumed to exist for t treatments in the uth of g groups. For the uth group the probability that treatment i obtains top ranking in the presence of treatments j and k is $\pi_{iu}/(\pi_{iu} + \pi_{ju} + \pi_{ku})$ and the probability that treatment j obtains rank 2, given that i had rank 1, is $\pi_{ju}/(\pi_{ju} + \pi_{ku})$.

The three test of hypotheses listed in the first paper are again developed. Tables are under preparation but are not yet available or complete.

6. **Limit Theorems Associated with Variants of the von Mises Statistic.** M. Rosenblatt, University of Chicago.

A multidimensional analogue of the von Mises statistic is considered for the case of sampling from a multidimensional uniform distribution. The limiting distribution of the statistic is shown to be that of a weighted sum of independent chi-square random variables with one degree of freedom. The weights are the eigenvalues of a positive definite symmetric function. A modified statistic of the von Mises type useful in setting up a two-sample test is shown to have the same limiting distribution under the null hypothesis (both samples come from the same population with a continuous distribution function) as that of the one-dimensional von Mises statistic. The paper makes use of elements of the theory of stochastic processes.

7. **A Modification of Schwarz's Inequality with Applications to Distributions.** Sigeiti Moriguti, University of North Carolina and University of Tokyo.

Let $\Phi(t)$ be a function of bounded variation in the closed interval $[a, b]$ and continuous at both ends. Then for any nondecreasing function $x(t)$ belonging to $L_2(a, b)$, and summable

with respect to Φ, $\int_a^b x(t)\, d\Phi(t) \leq \left\{\int_a^b x(t)^2\, dt\right\}^{\frac{1}{2}} \left\{\int_a^b \bar{\phi}(t)^2\, dt\right\}^{\frac{1}{2}}$, where $\bar{\phi}(t)$ is the right-hand derivative of the "greatest convex minorant" of $\Phi(t)$. This is proved and necessary and sufficient conditions for the equality to hold are also given. Several examples of application to distribution problems in statistics are discussed.

8. **Confidence Intervals of Fixed Geometric Size for Scale Parameters.** (Preliminary Report.) LIONEL WEISS, University of Virginia.

A procedure is given for obtaining confidence intervals for parameters of scale with confidence coefficient no less than β and length no greater than Δ, where β is any number between 0 and 1 and Δ is any positive number. The procedure uses two samples, the size of the second sample being a chance variable. It seems certain that there are other procedures for the same purpose yielding a smaller expected number of observations, but even in using the method given the problem of fixing the size of the first sample to minimize the expected number of observations is tedious computationally. A comparison is suggested between the expected number of observations and the number of observations required when an upper bound for the scale parameter is known and a single sample is used to get a confidence interval of at least a given confidence coefficient and of length bounded by a given number.

9. **On Lower Bounds of Powers of Certain Multivariate Tests.** S. N. ROY, University of North Carolina.

For multivariate normal populations tests of hypotheses were earlier offered for (i) equality of two covariance matrices; (ii) independence of two set of variates, and (iii) the analysis of variance situation. Lower bounds of the powers of such tests are now discussed. Here, for simplicity, under (iii) is considered the hypothesis of equality of respective means for k p-variate populations with a common covariance matrix Σ_2. Let S_1 denote the "covariance matrix of the sample means," S_2 the "pooled covariance matrix of sample error," Σ_1 the corresponding population matrix of means, H_0 the hypothesis (iii), and H an alternative. Then the critical region of the test at a level α is: $\theta_q \geq \theta_0$, where θ_0 is given by $P(\theta_q \geq \theta_0 \mid H_0) = \alpha$ and θ_q is the largest characteristic root of the matrix $S_1S_2^{-1}$ (positive semidefinite of rank $q \equiv \min(p, k - 1)$, a.e.). For the power we have the following lower bound:

$$P(\theta_q \geq \theta_0 \mid H) > 1 - \prod_{i=1}^{q} \{1 - P\,(\text{noncentral } F \geq \theta_0 \mid \Theta_i)\},$$

the noncentral F being with d.f. $(k - 1)$ and $(N - k)$ (N: total number of observations), and Θ_i's being the characteristic roots of the matrix $\Sigma_1\Sigma_2^{-1}$ (positive semidefinite of rank, say, $s \leq q$). Similar lower bounds are also readily available for (i) and (ii).

10. **Normal Multivariate Analysis and the Orthogonal Group.** A. T. JAMES, Princeton University.

The relationship of the orthogonal group, and its two coset spaces, the Grassmann and Stiefel manifolds, to normal multivariate sampling theory is discussed. The use of the Blaschke differential forms to represent the invariant measures on the two manifolds is illustrated by a derivation of the well known distribution of the canonical correlation coefficients in the null case. The distribution of n independent samples from a normal k-variate population is transformed into 3 independent distributions, viz., (a) essentially the Wishart distribution; (b) the distribution of the linear subspace spanned by the sample when

represented as k vectors in n-space; this is given by the invariant measure in the Grassmann manifold; (c) the invariant distribution of a $k \times k$ orthogonal matrix which determines the orientation of the k vectors in the k-dimensional linear subspace.

11. Exact Formulae in Sequential Analysis for Exponential Distributions. Johan H. B. Kemperman, Purdue University.

Let $a > 0$ and $b > 0$. Let $X_1, X_2, \cdots$ be a sequence of independent random variables with a common distribution (we assume $Pr(X_i \neq 0) > 0$). Put $Z_n = X_1 + X_2 + \cdots X_n$ and let N be the random variable which takes the value n if $-a < Z_k < b (k = 1, \cdots, n-1)$ and $Z_n \geqq b$ or $Z_n \leqq -a$. We put $\rho_n = Pr(N = n)$, $p_n = Pr(N = n, Z_n \leqq -a)$ and $q_n = Pr(N = n, Z_n \geqq b)$. Let D be an open connected region in the complex z-plane containing an interval G on the imaginary axis. We suppose that there exists a function $\psi(t)$ which is analytic in D and which in G takes the value $\phi(t) = E(e^{tX})$. Then, the function which for t in G is defined by $r_n(t) = \rho_n E(e^{tZ_n} \mid N = n)$ can be extended to an analytic function $r_n(t)$ in D. Moreover, there exists a constant $\theta(0 < \theta < 1)$ such that for each value t in D with $|\phi(t)| \geqq \theta$ we have $\Sigma_1^\infty r_n(t) (t)^{-n} = 1$ (Wald's fundamental identity). For the same values t, this relation may be differentiated term by term with respect to t. This generalization is used to obtain generating functions for p_n and q_n under certain conditions.

12. A Note on a Generalized Behrens-Fisher Problem. Henry Scheffé, Columbia University.

An exact solution [Henry Scheffé, "On solutions of the Behrens-Fisher problem, based on the t-distribution," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 35–44; "A note on the Behrens-Fisher problem," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 430–432] of the Behrens-Fisher problem, based on the t-distribution, is generalized to yield confidence intervals for a linear combination of unknown parameters.

13. Large-Sample Confidence Intervals for Density Function Values at Percentage Points. John E. Walsh, China Lake, California.

Let us consider a sample of size n from a population with density function $f(x)$. Let θ_p represent the $100p\%$ point of this population. A class of "well behaved" density functions is defined. This class seems to contain density functions which are capable of approximating most practical situations of a continuous type for $.05 \leqq p \leqq .95$. This paper presents some approximate confidence intervals for $f(\theta_p)$ for the case where $.05 \leqq p \leqq .95$ and the density function is of the "well behaved" class. These results hold for values of n which are only moderately large. The exact value of a confidence coefficient is not known but is determined within reasonably close limits. An approximate expression is obtained for deciding when n is sufficiently large for application of these results. The minimum sample sizes required depend on p and the confidence coefficient; they range from around fifty to several thousand. The confidence intervals are based on statistics of the form $x[(p + \epsilon)n + C\sqrt{n}] - x[(p - \epsilon) - C\sqrt{n}]$, where $x[z] = x$[integer nearest z] and $x[1], \cdots, x[n]$ are the sample values arranged in increasing order of magnitude. The quantity ϵ is a small but fixed number depending on p, while C is chosen so that a confidence interval of the desired order of magnitude is obtained.

14. Sequential Sufficient Statistics. R. R. Bahadur, Delhi, India.

The author defines sequential sufficiency and gives some characterizations of it. Let $x_1, x_2, \cdots$ be a sequence of abstract chance variables having a joint distribution p be-

longing to a family P of probability distributions. For each m let $X_{(m)}$ be the space of all points $(x_1, x_2, \cdots, x_m)$, and let t_m be a function on $X_{(m)}$ with arbitrary range such that t_m is a sufficient statistic for P when the sample space is $X_{(m)}$. Then $(t_1, t_2, \cdots)$ is said to be a sequential sufficient statistic if for any event A depending only on $x_1, x_2, \cdots$ and x_m the conditional probability of A given t_{m+1} equals the conditional expectation given t_{m+1} of the conditional probability of A given t_m, $(m = 1, 2, \cdots)$. The role of sequential sufficient statistics in sequential decision problems has been described elsewhere [RAGHU RAJ BAHADUR, "On sufficiency and statistical decision functions," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 609–610 (abstract)]. The main result established here is the following. If $x_1, x_2, \cdots$ and x_m are independently distributed and their joint distribution is absolutely continuous with respect to a fixed σ-finite measure λ_m, $(p \in P;\ m = 1, 2, \cdots)$, then $(t_1, t_2, \cdots)$ is a sequential sufficient statistic.

15. **Some Powerful Rank Order Tests.** WASSILY HOEFFDING, University of North Carolina.

It is shown that in certain cases there exist nonparametric tests which depend only on the ranks of the observations and whose power is arbitrarily close to the power of a standard parametric test if the sample is sufficiently large. For example, let $(x_1, y_1), \cdots, (x_n, y_n)$ be a random sample from a continuous bivariate distribution. Let H be the hypothesis that x and y are independent. Let r_i and s_i be the respective ranks of x_i and y_i. Let $h_n(k)$ be the expected value of the kth order statistic in a sample of n observations from a normal $(0, 1)$ distribution. Let $c_n = \sum_{i=1}^{n} h_n(r_i)h_n(s_i)$. Let k_n be the smallest number for which the probability of $|c_n| > k_n$ does not exceed α when H is true. Suppose that (x, y) has a bivariate normal distribution with correlation ρ (which may depend on n), and that the power of the standard product-moment correlation test of size α tends to a constant $\beta \leqq 1$ as $n \to \infty$. Then the power of the test which rejects H if $|c_n| > k_n$ tends to the same limit β. Similar results hold for two-sample tests, analysis of variance tests, etc. (Work sponsored by the Office of Naval Research.)

16. **Confidence Bounds for a Set of Means.** D. A. S. FRASER, University of Toronto.

The following problem was suggested to the author by Professor John Tukey: given $x_1, \cdots, x_n$ are normal and independent with means $\mu_1, \cdots, \mu_n$ and variance σ^2, to find an upper confidence bound (or confidence interval) for the set of means $\mu_1, \cdots, \mu_n$. This paper proves that, subject to mild restrictions on the type of bound, exact β-level confidence bounds (or intervals) do not exist (unless $n = 1$ or $\beta = 0, 1$). Incidental to the proof, bounds are obtained having at least β confidence: they are max $x_i + \lambda_{1-\beta}\sigma$ for the upper bound and (min $x_i + \lambda_{1-\frac{1}{2}(1-\beta)}\sigma$, max $x_i + \lambda_{\frac{1}{2}(1-\beta)}\sigma$) for the interval, where λ_α is the value exceeded with probability α by a standardized normal variate.

If the μ's are values of the location parameter for a distribution with density $f_\mu(x) = f(x - \mu)$, then a bound (interval) with at least β confidence is obtained by using the above formulas with $\sigma = 1$ and with α defined as the α point of the distribution having $\mu = 0$. If this class of distributions is bounded complete with respect to the location parameter μ (using at least all μ less than, say, zero), then exact upper bounds to not exist.

NEWS AND NOTICES

Readers are invited to submit to the secretary of the Institute news items of interest

Personal Items

Dr. Carl B. Allendoerfer, formerly Professor of Mathematics at Haverford College, Pennsylvania, has accepted an appointment as Professor and Executive Officer of the Mathematics Department, University of Washington, Seattle.

Mr. Ishver S. Bangdiwala, who came from India in 1950 to the University of North Carolina to do graduate work in statistics, has now been appointed as Assistant Consulting Statistician at the Agricultural Experiment Station, University of Puerto Rico, Rio Piedras, Puerto Rico.

Dr. Archie Blake has accepted a position as Head of the IBM Computing Section, Cornell Aeronautical Laboratory, Buffalo, New York.

Mr. K. A. Brownlee, formerly Chief, Test Design Branch, Plans and Evaluation Office, Dugway Proving Ground, Tooele, Utah, has joined the staff of the Committee on Statistics, University of Chicago, as an Assistant Professor. Mr. Brownlee's work will be principally in the Committee's Statistical Research Center.

Dr. Enrique Cansado, Professor at the University of Madrid and Chief of the Methodological Section of the Instituto Nacional de Estadística, Spain, has accepted a visiting assistant professorship in the Department of Mathematics, University of California, Los Angeles, to teach Stochastic Processes and Calculus of Probability for the academic year 1951–1952. During the previous academic year he held a Del Amo Foundation fellowship for research and studies in the U. S. A.

Dr. W. R. Church, Professor of Mathematics at the United States Naval Postgraduate School, has moved from Annapolis to Monterey, California. The Postgraduate School, which was founded at Annapolis and was a department of the U. S. Naval Academy, has been moved to Monterey.

Dr. C. W. Cotterman, formerly Associate Geneticist at the Heredity Clinic, University of Michigan, has joined the staff of the School of Veterinary Medicine, University of California, Davis.

Dr. Edgar P. King, who received his doctor's degree in mathematics in June, 1951, from the Carnegie Institute of Technology, is now employed as a mathematical statistician in the Statistical Engineering Laboratory, National Bureau of Standards, Washington D. C.

Dr. Kenneth H. Kramer, who received his doctor's degree in mathematics in February, 1952, from the Carnegie Institute of Technology, is now employed by the Youngstown Sheet and Tube Company, Youngstown, Ohio, as a development engineer on the staff of the Director of Mill Research and Development. In this job, he will serve in a consulting capacity on general statistical applications to quality control, production control and cost control.

Dr. R. A. Leibler of the Armed Forces Security Agency has accepted a position as mathematician at the Sandia Corporation, Albuquerque, New Mexico.

Dr. Ardie Lubin has left his position as Lecturer in Statistical Psychology at the Institute of Psychiatry, University of London, and is now a Research Psychologist with the Personnel Research Section, Adjutant-General's Office, U. S. Department of the Army.

Mr. John W. Morse, formerly Chief, Epidemologic Studies Section, Venereal Disease Division, Federal Security Building, Washington, D. C., has been transferred to Chile as technical adviser in the field of health and vital statistics. The work is being sponsored by the United States TCA Point IV Program.

Mr. James A. Pierce, who has been doing graduate work in mathematics and acting as Graduate Assistant at Purdue University, is now employed by Consolidated Vultee Aircraft Corporation as an Aerophysics Engineer in an Armament Evaluation and Vulnerability Section of the Technical Design Department.

Dr. William J. Schull has accepted a position as Junior Geneticist at the Heredity Clinic, University of Michigan. He has spent the past two years in Japan as a statistical consultant with the Atomic Bomb Casualty Commission studying the effects of atomic radiation.

Dwarka Nath Nanda

Dwarka Nath Nanda died in Delhi, India, on March 9, 1952. He was the author of three contributions to multivariate statistical analysis, published in the *Annals of Mathematical Statistics* in 1948 and 1950.

Dr. Nanda was born on April 11, 1916, received the B.Sc. and M.A. degrees in Mathematics from the University of Agra, served as Statistician in the Punjab Agricultural Department for five and a half years and in the Imperial Council of Agricultural Research for two years and became Director of Statistics for the state of Mayurbhanj. He studied two years at the University of North Carolina and received the Ph.D. degree there in Mathematical Statistics in 1948.

On return from North Carolina he was appointed Assistant Professor of Statistics at the Indian Council of Agricultural Research. He relinquished this post in May, 1949, to take up the appointment of Senior Scientific Officer, Statistics, at the Technical Development Establishment Laboratories, Kanpur, under the Ministry of Defence, Government of India.

He is survived by his wife and three children.

Wald Memorial Fund

As a tribute in memory of the late Professor Abraham Wald, who was killed with Mrs. Wald in an airplane accident in India December 13, 1950, a group of friends and colleagues are establishing a fund to help in defraying the expenses of a college education for his two children, Betty, eight years old, and Bobby, now four.

Trustees of the fund are Theodore W. Anderson, Howard Levene, and Mortimer Spiegelman. Contributions may be made payable to WALD MEMORIAL FUND and may be sent to Howard Levene, Box 23 Fayerweather, Columbia University, New York 27, New York.

New Secretary-Treasurer

Professor K. J. Arnold of the University of Wisconsin has been elected by the Council of the Institute of Mathematical Statistics as Secretary-Treasurer for the term July 1, 1952, to June 30, 1955. He has also been appointed to an associate professorship at Michigan State College. The new business office of the Institute will be at the Department of Mathematics, Michigan State College, East Lansing, Michigan, during Professor Arnold's term of office. Professor Arnold will continue to reside in Madison, Wisconsin, until the end of August, and will receive mail during July and August at North Hall, University of Wisconsin, Madison 6, Wisconsin.

Preparation of Problem and Source Materials for the Mathematical Training of Social Scientists

As readers of the *Annals* probably know, a Committee on the Mathematical Training of Social Scientists has been at work for some time. The Committee includes representatives from the following associations and societies: American Anthropological Association, American Economics Association, American Educational Research Association, American Farm Economics Association, American Political Science Association, American Psychological Association, American Sociological Society, American Statistical Association, Econometric Society, Institute of Mathematical Statistics, Mathematical Association of America, and Psychometric Society.

As the result of a suggestion from this Committee, the Social Science Research Council is now sponsoring a small group to work during the summer of 1952 at Dartmouth College, Hanover, N. H. This group will attempt to compile from the literature of the various social sciences lists of problems, extracts from sources, and references to sources that illustrate varieties of uses of mathematics in the social sciences. These compilations are expected to serve a number of important ends—e.g., to provide mathematicians with material for use in texts and courses designed for social scientists, to indicate the general dimensions of the mathematical training appropriate for students of the social sciences now and in the future, and to facilitate the study of mathematics by social scientists for whom organized courses are not available.

This Committee believes that the group referred to would find it most helpful if it could have a wide variety of suggestions from the various areas concerned. A general appeal for such suggestions is hereby made. They should be sent to Professor William G. Madow, Chairman, Committee on the Mathematical Training of Social Scientists, Baker Library, Hanover, N. H., up to August 15, and thereafter University of Illinois, Urbana, Ill.

Although the Committee does not wish to limit the suggestions to specific types of material, it would prefer greater emphasis on materials relating to the use of mathematics in the social sciences themselves than on those relating to

statistics, since the materials necessary for statistics are better known. Moreover, the Committee would suggest that those who respond not concern themselves with questions of duplication of what others would say, but give as much information as possible. This first request for assistance is aimed at providing those who are interested in this subject with an opportunity to make their views known to the Committee in as general terms as they wish.

Finally, the Committee would appreciate learning where programs of mathematical training intended for social scientists are now in existence or in process of development, and where mathematics at the level of the calculus or higher is required for undergraduate or graduate degrees in the social sciences or may be substituted for another requirement for a degree in a social science.

Statistical Summer Session, July 29 to August 15, 1952

The Department of Statistics and the Statistical Laboratory in cooperation with the Department of Mathematics and the Department of Industrial Engineering of the Virginia Polytechnic Institute will conduct a special statistical summer session, July 29 to August 15, 1952. The program will be for graduate students, research workers, and technicians in government and industry. Special offerings will be given in the statistics of taste testing, bio-assay, sampling, and engineering research and production. For further details write the Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.

Summer Seminar in Statistics

The third meeting of the Summer Seminar in Statistics will take place on the campus of the University of Connecticut, Storrs, Connecticut, during the three weeks of August 4–22, 1952. There will be one or two seminar sessions each day and a clinic on the treatment of problems in applications.

The first week, August 4–8, which will be devoted to the modifications of statistical techniques appropriate for chemistry, is being organized by Cuthbert Daniel and W. L. Gore. The second week, August 11–15, will be divided into two parts. The latter part, devoted to applications of minimax techniques, is being organized by J. L. Hodges. The third week, August 18–22, will be divided into two parts. The first part will be devoted to follow-up studies as they arise in medicine; this is being organized by Irwin Bross. The second part will be devoted to applications in actuarial work; this is being organized by Mortimer Spiegelman. Professor R. A. Fisher will be a member of the seminar during the first two weeks.

Those interested in the subjects under discussion are invited to attend by the day, week or other period. (A nominal registration fee will be collected.) For further information on reservations for campus housing, write to the Secretary of the Seminar, Professor D. F. Votaw, Jr., 210 Leet Oliver Memorial Hall, Yale University, New Haven, Connecticut. Suggestions of problems which might be presented before the clinic may also be sent to Professor Votaw.

Doctoral Dissertations in Statistics, 1951

Listed below are the doctorates conferred during the year 1951 in the United States and Canada for which the dissertations were written on topics in statistics (or for a degree in statistics). The university, month in which degree was conferred, major subject, minor subject, and the title of the dissertation are given in each case if available.

Helen Abbey (Doctor of Science in Hygiene), Johns Hopkins, June, major in biostatistics, "An Examination of the Reed-Frost Theory of Epidemics."

R. E. Bechhofer, Columbia, June, major in mathematical statistics, minor in statistical quality control, "The Effect of Preliminary Tests of Significance on the Size and Power of Certain Tests of Univariate Linear Hypotheses."

W. S. Connor, Jr., North Carolina, August, major in mathematical statistics, minor in economics, "The Structure of Balanced Incomplete Block Designs and the Impossibility of Certain Unsymmetrical Cases."

A. M. Dutton, Iowa State College, June, major in statistics, minor in genetics, "Statistical Analysis of Long-Term Agricultural Experiments."

L. A. Goodman, Princeton, October, major in mathematics, "The Estimation of Population Size Using Sequential Sampling Tagging Methods."

E. R. Immel, California, June, major in mathematics, "Problems of Estimation and of Hypothesis Testing Connected with Birth-and-Death Markov Processes."

G. B. Kallianpur, North Carolina, August, major in mathematical statistics, minor in mathematics, "Some Topics in the Theory of Stochastic Processes."

E. P. King, Carnegie Institute of Technology, June, major in mathematics, "The Operating Characteristic of the Control Chart for Sample Means when Process Standards Are Unspecified."

K. H. Kramer, Carnegie Institute of Technology, June, major in mathematics, "The Distribution of Range in Compositions of Normal Universes."

A. S. Littell (Doctor of Science in Public Health), Johns Hopkins, June, major in biostatistics, "Estimation of the T-Year Survival Rate from Follow-up Studies over a Limited Period of Time."

Paul Meier, Princeton, October, major in mathematics, "Weighted Means and Lattice Designs."

R. B. Murphy, Princeton, October, major in mathematics, "On Tests for Outlying Observations."

Ingram Olkin, North Carolina, June, major in mathematical statistics, minor in mathematics, "On Distribution Problems in Multivariate Analysis."

D. B. Owen, University of Washington, March, major in mathematics, "A Two Sample Test Procedure."

M. P. Peisakoff, Princeton, October, major in mathematics, "Transformation Parameters."

D. D. Rippe, Michigan, June, "Statistical Rank and Sampling Variation of the Results of Factorization of Covariance Matrices."

Milton Sobel, Columbia, January, major in mathematical statistics, minor in mathematics, "An Essentially Complete Class of Decision Functions for Certain Standard Sequential Problems."

W. F. Taylor, California, January, "On Tests of Hypotheses and Best Asymptotically Normal Estimates Related to Certain Biological Tests."

M. E. Terry, North Carolina, June, major in mathematical statistics, minor in experimental statistics and mathematics, "Some Rank Order Tests which are Most Powerful Against Specific Parametric Alternatives."

F. H. Tingey, University of Washington, August, major in mathematics, minor in meteorology, "Extension of Kolmogoroff Statistic to More Than One Dimension."

New Members

The following persons have been elected to membership in the Institute

(December 2, 1951 to March 1, 1952)

Boretti, Lodovico, Ph.D. (Univ. of Genoa), Assistant Professor of Mathematical Statistics, Institute of Statistics, University of Genoa, Italy.

Borsting, Jack R., B.A. (Oregon State College, Corvallis), Graduate Assistant, Department of Mathematics, University of Oregon, Eugene, Oregon.

Bowman, John R., Ph.D. (Univ. of Pittsburgh), Head, Department of Research in Physical Chemistry, Mellon Institute, Pittsburgh 13, Pennsylvania.

Bush, Kenneth A., Ph.D. (Univ. of North Carolina), Associate Professor, Department of Mathematics, State University of New York, Champlain College, Plattsburgh, New York.

Busk, Thoger, B.Sc. (Univ. of Copenhagen), Statistician, W. H. O. Tuberculosis Research Office, Copenhagen, *℅ Dr. Tvaergade 41, III, Copenhagen, K., Denmark.*

Ehrenfeld, Sylvain, A.M. (Columbia Univ.), Graduate student, Department of Mathematical Statistics, Columbia University, *370 Columbus Avenue, New York 24, New York.*

Elkin, William F., M.S. (Univ. of Michigan), Graduate Research Assistant, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina.

Gomberg, Louis, B.A. (New York University), Graduate Student, Department of Mathematical Statistics, Columbia University, *2120 Mapes Avenue, New York 60, New York.*

Green, William K., M.A. (Univ. of Illinois), Analyst, Department of Defense, Armed Forces Security Agency, Washington, D. C., *Apartment 102, 6616 Willston Place, Falls Church, Virginia.*

Hagan, John S., M.S. (St. Louis Univ.), Graduate student, Department of Mathematical Statistics, Columbia University, New York 27, New York, *4426 Broadway, Kansas City 2, Missouri.*

Heimlich, C. Roger, Student, Laboratory Instructor for Machine Methods, Purdue University, *726 N. Chauncey Avenue, West Lafayette, Indiana.*

Ito, Koichi, B.E. (Univ. of Tokyo), Graduate student, Department of Mathematics, Saint Louis University, *#83 Vandeventer Place, St. Louis 8, Missouri.*

Jones, Wayne H., M.S. (Univ. of Chicago), Analytical Statistician, Personnel Research Section, Personnel Research and Procedures Branch, Adjutant General's Office, Department of the Army, Washington, D. C., *915 Cofer Road, Falls Church, Virginia.*

Kozelka, Robert M., M.A. (Univ. of Minnesota), Graduate student, Department of Mathematics, Harvard University, *16 Usher Road, W. Medford 55, Massachusetts.*

Lamke, Tom A., Ph.D. (Univ. of Wisconsin), Research Specialist, Bureau of Research, Iowa State Teachers College, Cedar Falls, Iowa.

Lehrer, Thomas A., M.A. (Harvard), Graduate student, Department of Mathematics, Harvard University, *6 Kirkland Road, Cambridge 38, Massachusetts.*

Lindquist, E. F., Ph.D. (State Univ. of Iowa), Professor of Education, College of Education, State University of Iowa, Iowa City, Iowa.

Miller, Raphael, M.A. (Yale), Graduate student, Department of Mathematics, Yale University, *104 York Square, New Haven, Connecticut.*

Miyashita, Totaro, M.A. (Univ. of Tokyo), Graduate student, Department of Economics, University of Chicago, *Room 472, International House, 1414 East 59th Street, Chicago 37, Illinois.*

Moore, Roger H., B.S. (Univ. of Oregon), Research Assistant and Graduate student, Department of Mathematics, University of Oregon, *1319 East 15th Street, Eugene, Oregon.*

Newell, Charles R., B.Sc. (Univ. of Toronto), Statistician, Norton Company, Chippawa, Ontario, *1761 Peer Street, Niagara Falls, Ontario, Canada.*

Nicholson, Wesley L., A.B. (Univ. of Oregon), Research Assistant and Graduate student, Department of Mathematics, University of Oregon, *446 E. 13th Street, Eugene, Oregon.*

North, John D., Chairman and Managing Director, Boulton Paul Aircraft, Ltd., Wolverhampton, England.

Palmer, Boyd Z., B.A. (Earlham College, Richmond, Indiana), Graduate student, Department of Mathematical Statistics, University of North Carolina, *University Trailer Court, #95, Chapel Hill, North Carolina.*

Romero, Mario G., Student, George Washington University, Washington, D. C. and employee of Section on Mathematical Statistics, Bureau of Statistics, Direccion General de Estadística, San Jose, Costa Rica, Central America.

Rosenblatt, Harry M., B.S. (George Washington Univ.), Graduate student, Department of Mathematical Statistics, George Washington University, and Mathematical Statistician, Navy Department, Bureau of Ordnance, *506 Oglethorpe St., N. E., Washington D. C.*

Schmid, John, Jr., Ph.D. (Univ. of Wisconsin), Assistant Professor and Examiner, Board of Examiners, Room 5, Berkey Hall, Michigan State College, East Lansing, Michigan.

Tsao, Chia K., M.A. (Univ. of Oregon), Graduate Assistant, Statistical Laboratory, and graduate student, Department of Mathematics, University of Oregon, Eugene, Oregon.

Tucker, Howard G., M.A. (Univ. of Calif.), Graduate student, Department of Mathematical Statistics, University of California, *1074 Spruce Street, Berkeley 7, California.*

Uemura, Kazuo, Kogakushi (Univ. of Tokyo), Graduate student, Department of Mathematical Statistics, University of North Carolina, *226 B. Dorm, Chapel Hill, North Carolina.*

Walker, Andrew M., M.A. (Univ. of Cambridge), Research Assistant in Mathematical Statistics, Mathematics Department, The University, Manchester 13, England.

Winston, Gerald, M.A. (Columbia Univ.), Chief Statistician, Research and Development Branch, Philadelphia Quartermaster Depot, 2800 S. 20th St., Philadelphia 45, Pennsylvania.

Zahl, Samuel, B.S. (Univ. of Chicago), Graduate student, Department of Mathematics, University of Chicago, *4340 S. Drexel Boulevard, Chicago, Illinois.*

REPORT OF THE BLACKSBURG MEETING OF THE INSTITUTE

The fifty-first meeting of the Institute of Mathematical Statistics was held at the Virginia Polytechnic Institute, Blacksburg, Virginia on March 19–21, 1952, with the Biometric Society (Eastern North American Region). Ninety persons attended the meeting, including the following thirty-seven members of the Institute:

R. L. Anderson, R. E. Bechhofer, Z. W. Birnbaum, R. C. Bose, R. A. Bradley, A. E. Brandt, G. L. Burrows, E. L. Cox, Gertrude Cox, B. de Loor, D. B. Duncan, Churchill Eisenhart, R. E. Greenwood, R. J. Hader, Boyd Harshbarger, Harold Hotelling, Paul Irick, A. T. James, A. W. Kimball, C. F. Kossack, T. E. Kurtz, R. F. Link, Paul Meier, Sigeiti Moriguti, M. H. Quenouille, S. N. Roy, Henry Scheffé, S. A. Schmitt, H. Fairfield Smith, Harry Smith, Jr., Henry Teicher, M. E. Terry, J. W. Tukey, G. W. Tyler, Kazuo Uemura, D. L. Wallace, Lionel Weiss.

At the opening session, Wednesday morning, March 19, Professor M. H. Quenouille, Yale University, gave an address entitled *The Consequences of Testing Significance*. Professor J. M. Grayson, Virginia Polytechnic Institute, was the chairman and H. Fairfield Smith, North Carolina State College, was a discussant.

The second session, held Wednesday afternoon, at which Professor C. F. Kossack, Purdue University, was chairman, was devoted to *Multiple Comparisons*. At this session the following papers were given:

1. *On a Multiple Decision Procedure Associated with Certain Ranking Problems.* Robert E. Bechhofer, Columbia University.
2. *The Multiple Comparisons Test for Separating Ranked Treatments in an Analysis of Variance.* D. B. Duncan, Virginia Polytechnic Institute.
3. *Allowances for Various Types of Error Rates.* John W. Tukey, Princeton University.
4. *Short Cuts to Allowances.* R. F. Link and D. L. Wallace, Princeton University.

The Thursday morning session was devoted to contributed papers of the Biometric Society.

At 2:00 P.M., Thursday, Professor Robert J. Hader, North Carolina State College, gave an address on *Double Sampling Acceptance Inspection on Measurable Quality Characteristics.* Professor H. L. Manning, Virginia Polytechnic Institute, was chairman and Professor R. C. Bose, University of North Carolina, was a discussant.

At 3:30 P.M., Thursday, a joint session for contributed papers of the Institute and the Biometric Society was held with Professor P. S. Dear, Virginia Polytechnic Institute, as chairman. The following papers were presented:

1. *On the Approximation of Sampling Distributions by Punch Card Methods.* Carl F. Kossack and Lester L. Helms, Purdue University.
2. *Resolvable Incomplete Block Designs with Two Replications.* R. C. Bose and K. R. Nair, University of North Carolina.
3. *Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons.* R. A. Bradley and M. E. Terry, Virginia Polytechnic Institute.

4. *Multiple Regression with a Quantal Response.* D. B. Duncan and R. C. Rhodes, Virginia Polytechnic Institute.
5. *Rank Analysis of Incomplete Block Designs. II. The Method for Blocks of Three. Preliminary Report.* (By title.) R. A. Bradley and M. E. Terry, Virginia Polytechnic Institute.

At 7:00 P.M., Thursday, the banquet was held, with Professor Boyd Harshbarger, Virginia Polytechnic Institute, acting as chairman. Dr. Louis A. Pardue, Vice President of Virginia Polytechnic Institute, gave a welcome. Dr. H. N. Young, Director, Virginia Agricultural Experiment Station, gave an address entitled *Administration of Research.*

The Friday morning session was devoted to contributed papers of the Institute. Professor R. A. Bradley, Virginia Polytechnic Institute, was chairman. The following papers were presented:

1. *Limit Theorems Associated with Variants of the von Mises Statistic.* M. Rosenblatt, University of Chicago.
2. *A Modification of Schwarz's Inequality with Applications to Distributions.* Sigeiti Moriguti, University of North Carolina and University of Tokyo.
3. *Confidence Intervals of Fixed Geometric Size for Scale Parameters. Preliminary Report.* (By title.) Lionel Weiss, University of Virginia.
4. *On Lower Bounds of Powers of Certain Multivariate Tests.* S. N. Roy, University of North Carolina.
5. *Normal Multivariate Analysis and the Orthogonal Group.* A. T. James, Princeton University.
6. *Exact Formulae in Sequential Analysis for Exponential Distributions.* Johan H. B. Kemperman, Purdue University.
7. *A Note on a Generalized Behrens-Fisher Problem.* (By title.) Henry Scheffé, Columbia University.
8. *Large-Sample Confidence Intervals for Density Function Values at Percentage Points.* (By title.) John E. Walsh, China Lake, California.
9. *Sequential Sufficient Statistics.* (By title.) R. R. Bahadur, Delhi, India.
10. *Some Powerful Rank Order Tests.* (By title.) Wassily Hoeffding, University of North Carolina.
11. *Confidence Bounds for a Set of Means.* (By title.) D. A. S. Fraser, University of Toronto.

At 2:00 P.M., Friday, Dr. John H. Curtiss, National Bureau of Standards, gave an address entitled *Some Chain Functions Useful in the Monte Carlo Method.* Dr. L. A. Pardue, Virginia Polytechnic Institute, was chairman.

At 3:15 P.M., Friday, a session was held on *Experimental Design*, at which Mr. Glenn Burrows, Bureau of Agricultural Economics, acted as chairman and Professor R. C. Bose, University of North Carolina, was a discussant. The following papers were presented:

1. *Recent Developments in Incomplete Block Designs.* K. R. Nair, Forestry Research Institute, Dehra Dun, India.
2. *Latinized Rectangular Lattices.* Boyd Harshbarger, Virginia Polytechnic Institute.

BOYD HARSHBARGER
Assistant Secretary

PUBLICATIONS RECEIVED

Francisco Quiroz Cuaron, *Tablas de Precios de Compra de Valores, Tomo I*, Banco de Mexico, S.A., Mexico City, 1952, 507 pp.

André G. Laurent, *La Méthode Statistique dans l'Industrie*, Presses Universitaires de France, Paris, 1950, 134 pp.

THE INSTITUTE OF [illegible]

[illegible]

The purpose [illegible] research in the [illegible] between the [illegible]

Membership [illegible] STATISTICS are [illegible] elsewhere. [illegible] should be sent to the [illegible]

MEETINGS [illegible]

[illegible]

ANNUAL MEETING, CHICAGO [illegible]